TO ERR AND ERR, BUT LESS AND LESS

Predictive coding and affective value in perception, art, and autism

SANDER VAN DE CRUYS

Dissertation offered to obtain the degree of
Doctor of Psychology (PhD)

SUPERVISOR:
Prof. Dr. Johan Wagemans

GESTALTREVISION
Perceptual organization in the context of a dynamical and hierarchical visual brain

Laboratory of Experimental Psychology
Faculty of Psychology and Pedagogical Sciences
KU Leuven

2014

Cover image created by RANDALL CASAER.

Title adapted from the poem *The road to wisdom* by the Danish scientist PIET HEIN.

Back cover image from *Les Songes Drolatiques de Pantagruel* (1565) attributed to FRANÇOIS DESPREZ, but based on the fabulous, chimerical characters invented by FRANÇOIS RABELAIS.

Typographical style based on *classicthesis* developed by ANDRÉ MIEDE (GNU General Public Licence). Typeset in LaTeX.

ABSTRACT

The importance of prediction or expectation in the functioning of the mind is appreciated at least since the birth of psychology as a separate discipline. From minute moment-to-moment predictions of the movements of a melody or a ball, to the long-term plans of our friends and foes, we continuously predict the world around us, because we learned the statistical regularities that govern it. It is often only when predictions go awry —when the sensory input does not match with the predictions we implicitly formed— that we become conscious of this incessant predictive activity of our brains. In the last decennia, a computational model called predictive coding emerged that attempts to formalize this matching process, hence explaining perceptual inference and learning. The predictive coding scheme describes how each level in the cortical processing hierarchy predicts inputs from levels below. In this way resources can be focused on that part of the input that was unpredicted, and therefore signals important changes in the environment that are still to be explained. In doing so the brain is guided by a single principle known as the prediction error minimization principle. Its constant effort to reduce uncertainty relative to its predictions is thought to enable adaptive behavior in a computationally manageable way.

The appeal of this account is clear from the recent surge in studies using it to explain findings across virtually all subfields of psychology. Indeed, a generalized form of predictive coding has been proposed as *Grand Unified Theory* for psychology. Unpacking the prediction error minimization principle in these disparate fields has led to powerful, integrating insights. But is this just another short-lived fad in psychology? Or are its roots deeper and its branches as far-reaching as its proponents foreshadow? We will only know if we examine the framework (its assumptions, implementations, implications and limitations) in full detail. The main mission of this dissertation is to do a small but, so we will argue, important part of this work. Specifically, we focus on how affective processes can have their place in this theory, since this has so far remained underexposed, given the mainly cognitive and perceptual character of previous work. To do so, we start with two elaborate 'case studies'. The first applies predictive coding to autism spectrum disorders, a psychiatric syndrome characterized by difficulties in social, cognitive and affective processing. We argue that the weight (or 'precision') attributed to prediction errors in autism is unduly high and inflexible, rather than determined by context uncertainty. This may explain both their peculiar behavior in cognitive and perceptual tasks, and their affective-motivational troubles. The second case is an application of predictive coding to visual art. Here we develop the hypothesis that making predictive progress (i. e., actively reducing prediction errors) is intrinsically pleasurable, which helps to explain affective experiences in art appreciation. These theoretical studies are followed by two empirical works, of which the first is an initial step in subjecting the hypothesis of predictive progress and pleasure to the test, using Gestalt discovery in two-tone images. The second investigates whether perception can be biased by affective relevance of the top-down priors or predictions applied to interpret a stimulus. It uses bistable point-light figures and finds some support that this is indeed the case. The closing chapter attempts to zoom out and build a coherent account of positive and negative affective value within the confines of the predictive coding framework. The core explanatory factor here is the rate of prediction error reduction and the (unexpected) changes in these rates. A host of evidence seems consistent with this idea that affective experience reflects a form of non-conceptual metacognition about prediction error dynamics.

In sum, this new application of predictive coding in issues of emotional value is worth following through, because of its parsimony, explanatory power, and the specific, testable research questions that can be derived from it. Most likely, the theory will be further refined by a joint effort of neurocomputational modelers and experimental psychologists, for which we set out the tracks in the different chapters.

SAMENVATTING

Psychologen onderkennen het belang van predicties of verwachtingen in het functioneren van de geest al sinds het ontstaan van hun discipline. Van de baan van een bal of het verloop van een melodie tot de plannen van onze partners en tegenstrevers, wij voorspellen voortdurend de wereld rondom ons, omdat we de statistische relaties die erin van kracht zijn geleerd hebben. Pas wanneer onze predicties mis zijn —wanneer de sensoriële informatie niet overeenstemt met de predicties die we er impliciet over vormen— worden we ons bewust van die voortdurende predictieve activiteit van onze hersenen. In de voorbije decennia ontwikkelden cognitieve neurowetenschappers een computationeel model met de naam *predictive coding*, dat dit iteratief matchen van predicties met input probeert te formaliseren, en zo perceptueel leren and inferentie tracht te verklaren. Het model beschrijft hoe elk niveau in de corticale hierarchie van perceptuele verwerking de neurale activiteit in een lager niveau voorspelt, zodat verwerkingsmiddelen toegespitst kunnen worden op het deel van de input dat onverwacht was. Dit deel omvat immers belangrijke veranderingen in de omgeving die nog niet opgenomen zijn in ons huidig mentaal model. De hersenen volgen hierbij één centraal principe bekend als het predictiefout-minimaliseringsprincipe. De constante inspanningen van het brein om de onzekerheid relatief tot zijn predicties te reduceren zouden adaptief gedrag mogelijk maken op een computationeel beheersbare manier.

Een recente groei in studies die predictive coding aanwenden om bevindingen te verklaren in praktisch alle subdomeinen van de psychologie, is een bewijs van de aantrekkingskracht van dit model. Meer nog, een veralgemeende vorm van predictive coding wordt naar voor geschoven als unificerende theorie voor psychologie en de neurowetenschappen. Het doordenken van het predictiefout-minimaliseringsprincipe in de uiteenlopende subdomeinen heeft geleid to krachtige, integrerende inzichten. Maar is dit simpelweg de nieuwe rage in de psychologie, intens maar snel weer vergeten? Of reiken de wortels ervan dieper en de vertakkingen zo ver als zijn verdedigers ons voorhouden? We weten het pas als we het model kritisch onderzocht hebben; zijn vooronderstellingen, implicaties en beperkingen inbegrepen. Het centrale doel van dit proefschrift is om van dat werk een klein maar, zo zullen we betogen, belangrijk deel te doen. We richten meer specifiek de aandacht op de plaats die affectieve processen kunnen hebben in dit kader, omdat dit tot nu toe onderbelicht bleef in het hoofdzakelijk cognitief en perceptueel getinte werk over predictive coding. We beginnen dit werk met twee uitvoerige 'gevalsstudies'. In de eerste passen we predictive coding toe op autismespectrumstoornis, een psychiatrische syndroom gekenmerkt door moeilijkheden in sociale, cognitieve en affectieve verwerking. We beargumenteren dat het gewicht dat toegekend wordt aan predictiefouten (technisch: de 'precisie') in autisme onterecht hoog en inflexibel is, in plaats van aangepast aan onzekerheid in de context. Vanuit dat gegeven kunnen we zowel ongewoon gedrag in cognitieve en perceptuele taken begrijpen, als de affectief-motivationele problemen. In de tweede casus passen we predictive coding toe op visuele kunst. Hier ontwikkelen we de hypothese dat predictieve vooruitgang boeken (d.i., het actief reduceren van predictiefouten) intrinsiek plezierig is, en dat dit onze affectieve beleving van kunst helpt verklaren. Na deze theoretische studies volgen twee empirische werken, waarvan het eerste poogt de hypothese over predictieve vooruitgang en positief affect aan de test te onderwerpen. We gebruiken daarbij ontdekking van Gestalt (voorspelbare structuur) in vervormde afbeeldingen. Het tweede empirische werk bekijkt of perceptie (van bistabiele puntlichtfiguren) vertekend kan zijn door affectieve relevantie van predicties die we hanteren om een stimulus te interpreteren. In het afsluitende hoofdstuk zoomen we uit en trachten we een coherente invulling te geven aan positief en negatief affect binnen het computationeel kader van predictive coding. De verklarende factor aan de basis hier is verandering in predictiefouten over tijd (technisch: de eerste afgeleide van predictiefout) en (onverwachte) veranderingen in die 'tempo's'. We bespreken een rist bevindingen die compatibel lijkt te zijn met het idee dat affectieve ervaring een weerspiegeling is van een vorm van niet-conceptuele metacognitie over predictiefoutdynamieken.

We besluiten dat deze nieuwe uitwerking van predictive coding in relatie tot kwesties over emotionele waarde het waard is om verder uitgespit te worden, omwille van haar spaarzaamheid, verklarende kracht, en de specifieke, toetsbare onderzoeksvragen die eruit voortvloeien. Naar alle waarschijnlijkheid zal de theorie verder verfijnd worden door de gezamenlijke inzet van neurocomputationele modelleerders en experimenteel psychologen, waarvoor we krijtlijnen uitzetten in de verschillende hoofdstukken.

We work in the dark —we do what we can— we give what we have.
Our doubt is our passion, and our passion is our task.
The rest is the madness of art.

— Henry James, 1893

Nature is very exact in the matter. It hurts just as much as it is worth.

— Julian Barnes

Dankwoord

Dit boekje was er ei zo na niet gekomen. Als het er dan toch ligt, is dat te danken aan een hele rist fantastische mensen, die ik hier met een diepe buiging wil aflopen.

Angst is, tegen wil en dank, altijd al een companion geweest. Een radicale twijfel, die nietsontziend tot op mijn bot gaat. Met het paranoïde besef dat wat je het meest in vraag moet stellen, het meest verborgen blijft voor jezelf. Dat neurotisch getob is in de wetenschap natuurlijk gemakkelijk om te buigen tot iets goeds. Tegelijk voed je daarmee het beest. Het vergroot zijn greep en kent zijn perken niet meer. Het denken dat zich tot zichzelf richt en daarmee soms een autofage ravage aanricht. Koud zweet en hete strijd. Dat moest ik tot in het diepst van mijn vezels ondervinden. Is de passie op de angst geënt, of andersom? Ik weet alleen dat ze bij mij als yin en yang samengaan. Maar het blijft een explosieve combinatie. De passie geeft een intellectuele gulzigheid (wie komt er op het hubristische idee om zowel empirisch als theoretisch onderzoek te willen doen over zo'n uiteenlopende materie?). De angst belet je alles te doen wat minder dan perfect is (veel te laat ontdekt dat het najagen van het ultieme experiment vruchteloos is). *Analysis paralysis*, of hoe iets vreselijks zo mooi kan klinken.

Een wetenschapper is een eeuwige beginneling, voortdurend botsend niet alleen tegen wat we, als mensen, weten en kunnen, maar ook tegen je eigen persoonlijke limieten. Zonder garantie dat die limieten slechts van tijdelijke aard zijn. Toch vind je ook troost in de wetenschap. In de zin dat, als het allemaal hoe dan ook eindigt in meer entropie, ik me nu wel een aantal gevechten kon besparen, toch? Of in een hervonden bevlogenheid om een onwijs interessante kwestie volledig uit te spitten. En de voldoening om er iets inzichtelijks over neer te kunnen schrijven. De rush van, net voor slapengaan, een paper te ontdekken die zo verschrikkelijk boeiend is, dat ie echt niet kan wachten tot de volgende ochtend (ok, middag).

Boven alles vind je troost in de ontdekking dat je medemensen ook in die gevechten verwikkeld zitten. Als ik mijn sisyphussteen dan toch definitief boven gekregen heb, is dat dus mede te danken aan mijn geweldige omgeving.

Johan die, geruggesteund door de Methusalemfinanciering, mij de ruimte gaf om mijn eigen neus te volgen, zelfs als die mij ver buiten het domein van de visuele perceptie leidde. Als ongeplande nevenprojecten uitgroeiden tot kolossen die soms moeilijk te beteugelen waren. Hoewel ik die ruimte aanvankelijk wel eens vervloekte, ben ik blij en fier over het pad dat ik erdoor heb kunnen afleggen. Dank, Johan, voor het geduld met mijn *slow science* ritme. Ik besef ten volle dat dit geen evidentie meer is te midden van de toenemende vermarkting van de wetenschappen. Het siert je enorm dat je de Methusalemfondsen op een zo pluralistische wijze inzet, en daarbij wat risico niet schuwt.

Hij mag dan wel een omvangrijk lab runnen, Johan is steeds bij de pinken om blinde vlekken weg te werken en de puntjes op de i te zetten. Maar bovenal apprecieer ik de vanzelfsprekendheid waarmee hij van in het prille begin in mij geloofd heeft. Hij drukte zo meer dan eens mijn *imposter* syndroom de kop in: de gevreesde mand met de fragiele bodem verdween zowaar even van de radar. Johan had waarschijnlijk snel in de smiezen dat ik op tijd en stond een forse duw nodig had om door te zetten, of een fikse aanmoediging als ik de kaas van mijn boterham dreigde te laten eten.

Ik wil ook mijn co-auteurs enorm bedanken. Ben heeft me wegwijs gemaakt in de wereld van de puntlichtventjes en in de wetenschappelijke best practices bij mijn eerste experimentele stappen. Els Nieuwlandt durfde het aan mijn thesisstudent te zijn en nam experimenten af als had ze nooit anders gedaan. Een bijzondere dankjewel ook aan de autisme-experten verbonden aan het lab: Lien, Bart, Ruth en Kris. Zonder hen had ik me nooit zo snel kunnen inwerken in hun domein. Het was een voorrecht om met hen samen te overleggen en schrijven. Kris en Ruth kan ik niet genoeg bewierooken: werkelijk topdames, niet alleen om wat ze doen, maar vooral ook om wie ze zijn. Bedankt voor de warme babbels! Ik wil verder ook Lee speciaal vermelden: hij wakkerde mijn enthousiasme over *predictive coding* aan, en deelt mijn fascinatie voor domein-overstijgende psychologische kwesties, wat voor ontiegelijk interessante discussies zorgde. Bedankt allen!

A huge thanks also to the members of the jury and of the midterm advisory committee: Gilles Pourtois, Elizabeth Pellicano, Tom Beckers, Peter Kuppens, Dirk Hermans, Raymond van Ee and René Zeelenberg. They dedicated valuable time to the reading and com-

menting of this work; I can only hope that it deserved that time and that it gave them some of the pleasure that I had when writing it.

I am much indebted to Jakob Hohwy and Karl Friston. Their work has been a tremendous inspiration, and both have given me invaluable feedback and generous encouragements at crucial moments in the development of my own thinking. I can't claim to understand all of the mathematical intricacies of Karl's work but, like a good work of art, it keeps me coming back for another nugget of insight.

To all colleagues that willingly read and constructively commented on earlier (sometimes lengthy) versions of one of the chapters before you: Raymond van Ee, John Brock, Floris de Lange, Vebjörn Ekroll, Marie Gomot, Jakob Hohwy, Ilona Kovacs, Laurent Mottron, Colin Palmer, Jean Steyaert, Jeroen van Boxtel, Peter van der Helm, Cees van Leeuwen, Michael Kubovy, Dorothee Augustin, and Neil Dodgson. Thanks a lot! You pushed my writings to a higher level. The presented views and all remaining shortcomings are, of course, solely my own responsibility.

Al die knappe labgenoten dan. Qua topics was ik misschien een buitenbeentje in het lab. Toch had dit werk niet in een ander lab geschreven kunnen zijn. Dat blijkt (voor de goede verstaander) uit elke bladzijde. Vaak zonder het te beseffen hebben ze mijn onderzoekspaden wezenlijk bepaald. Maar niet minder belangrijk: ze maken het bijzonder prettig vertoeven in het lab. De guitige maanden met Tom en Katrien behoren tot de memorabelste van de hele periode. Geen idee wat het was, waardoor we compleet *nuts* gingen, maar het was geweldig. Een welgemeende dankjewel. *Tusen takk* Vebjörn, voor je broederlijke, nuchtere raad en je lucide inzicht in de raderen van de academische wereld. Mark en Pieter heb ik ook erg hoog zitten. Mark doorprikt met één welgemikte, kurkdroge opmerking een situatie die veel te veel serieux opbouwde, wat bij mij steevast een gigantische zucht van *comic relief* uitlokt. Over Pieter zeiden Stephanie en ik al van toen hij de eerste keren opdook in het lab dat hij een grote meneer is. En toen ging het nog niet eens over zijn statistische *wizardry*, maar gewoon over zijn minzaam en bescheiden optreden. Ik wil ze eigenlijk gewoon allemaal vernoemen: Sven, Naoki, Nathalie, Bart, Becky, Bilge, Steven, Jonas, Maarten, Kathleen, Caitlin, Nihan, Krista, Chris, Tomer, Sasha, Erik, Radha, Nick: mijn appreciatie voor hen en wat ze doen is torenhoog.

Professor-provocateur Pol staat ook gebeiteld in mijn hoofd: de wiskundige steekvlieg die speciaal afdaalde van't stad om ons, ondermaanse psychologen, op tijd en stond wakker te schudden uit onze bedrijvige roes. Met wisselend succes overigens, maar met nietaflatende goesting.

Bedankt, Rudy, Agna en Marleen, ons immer presente, onvolprezen support team, om die warrige wetenschappers toch nog enige houvast te bieden. Rudy stond onder andere paraat om me te depanneren als de hond in de laatste maanden alsnog besloot om mijn doctoraat te saboteren door heel geslepen mijn levenslijn, aka laptopkabel, door te bijten. Beiden (de hond en Rudy) hebben ze er overigens op toegezien dat ik gezond bleef in de eindspurt, de eerste door middel van een frisse neus, de tweede met een frisse pint. Bedankt!

Stephanie, als ik het uiteindelijk toch gehaald heb, is het niet in het minst ook aan jou te danken. Geheel belangeloos nam je mij onder je vleugels tot ik de mijne weer zelf wat kon spreiden. Zoiets blijft je leven lang bij.

Tony 'badabing' Torfs, gewoon een prachtkerel. En goeie wijn behoeft geen krans. Bedankt, man. Tim, Joris en al die andere Kempense parels, de ontmoetingen zijn schaarser in de laatste jaren, maar des te waardevoller.

Mijn ouders, alledrie, die me mateloos steunen en toejuichen van de zijlijn. Mijn dankbaarheid aan hen is met geen pen te beschrijven. Ik durf te bevroeden dat ze zich wel eens voor het hoofd sloegen als ze te veel herkenbare zwartgalligheid in mijn ogen konden lezen. Ik bedenk me alleen hoeveel zwaarder het zou geweest zijn, met ouders die daar helemaal *geen* voeling mee hebben.

Mijn lieve broer en zus, Pieter en Roos, en hun respectievelijke eega, Deborah en Jeff. Ik kus mijn twee pollen voor zo'n ingoeie familie. Ze krijgen mij altijd terug op de rails, al dan niet door het inzetten van hun nog kleine, maar krachtige *weapons of mass distraction*. Noah, Lenthe en Axl, de probate antidota tegen een mineur gemoed. Dank jullie wel! Zo ook voor mijn schoonfamilie die me altijd met veel zorg en warmte omringen. Dankuzeer!

Ik hou natuurlijk het lekkerste tot het laatste: Mijn liefste, Lies. *You're so easy to love and you make loving me seem easy.* Bedankt om gedurende vele, veel te lange nachten mijn dwangmatig speur- en schrijfgedrag te tolereren. Zonder jou swingen onrust en verwarring mijn hersenpan uit. In één beweging stil je mijn trekken en demonen. Bedankt om me te ontzien in mijn sikkeneurige buien, of wanneer ik weer eens opgesloten zat in mezelf, of als huishoudelijke taken weer op jouw schouders terecht kwamen,... Ik beloof met de verbouwing van ons huis, hetzelfde met mezelve te proberen.

Dank, dank, dank.

Sander Van de Cruys
Leuven, 10 oktober 2014

Publications

These chapters (or some of the ideas) have previously appeared in the following publications:

- Van de Cruys, S., Wagemans, J. (2011). Gestalts as predictions: Some reflections and an application to art. *Gestalt Theory, 33*(3-4), 325-344. (Chapter 3)
- Van de Cruys, S., Wagemans, J. (2011). Putting reward in art: A tentative prediction error account of visual art. *i-Perception, (special issue on Art & Perception), 2*(9), 1035-1062. (Chapter 3)
- Van de Cruys, S., Schouten, B., Wagemans, J. (2013). An anxiety-induced bias in the perception of a bistable point-light walker. *Acta Psychologica, 144*(3), 548-553. (Chapter 5)
- Van de Cruys, S., de-Wit, L., Evers, K., Boets, B., Wagemans, J. (2013). Weak priors versus overfitting of predictions in autism: Reply to Pellicano and Burr (TICS 2012). *i-Perception, 4*(2), 95-97. (Chapter 2)
- Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de-Wit, L., Wagemans (2014). Precise minds in uncertain worlds: Predictive coding in autism. *Psychological Review*, in press. (Chapter 2)

IN PREPARATION

- Van de Cruys, S., Wagemans, J. (in preparation). Is Gestalt discovery intrinsically pleasurable? (Chapter 4)
- Van de Cruys, S. (in preparation). Value in the predictive brain. (Chapter 6)

Contents

List of Figures

List of Tables

Acronyms

ASD Autism Spectrum Disorders

EPF Enhanced Perceptual functioning

HIPPEA High, Inflexible Precision of Prediction Errors in Autism

PEM principle of Prediction Error Minimization

TD Typically Developing

WCC Weak Central Coherence

We demand rigidly defined areas of doubt and uncertainty!

— Douglas Adams, 1979

1

Introduction

Our relationship to uncertainty is fraught with contradictions. We can't live with or without it. Our resistance to uncertainty is grounded in our bodily existence, subject to the second law of thermodynamics: the law of increasing entropy. Continuous work is necessary to keep the organism in the far-from-equilibrium state that enables its very existence. In other words, the set of bodily states the organism can be in is constrained, because only these states are compatible with continued existence. This constraint on entropy of bodily states translates into informational constraints —when we do not know what is where, we cannot prevent the body from exceeding the viable states. Unexpected events can disturb our physiological integrity, so reducing uncertainty is an essential task for organisms to survive.

But incomplete knowledge, and hence uncertainty, is a fact of life. What is more, one could easily make the case that uncertainty is the spice of life. Indeed, without uncertainty, there is no striving, no motivation. Key characteristics of what it means to be alive would be missing. Eliminating uncertainty seems as close to death as to nirwana. In this work we will closely examine the link between uncertainty and striving or emotion in the broadest sense. To do so, we will take an increasingly supported model of perception as a starting point, because indeed, perception is thought to be about the mechanics of uncertainty reduction. However, while it is often considered to be disjointed from the rest of the living organism — an undirected and disinterested data crunching mechanism— we will see that this model allows a wider view grounded in our bodily existence. In fact, the model has been proposed as a *grand unified theory* of the mind-brain, including action, perception, learning and (social) cognition. A second main theme of the current work is to evaluate such claims in relation to issues of affective value and experience. The key criterion will be whether it allows one to formulate new, specific and interesting research questions, without losing explanatory power with regard to existing findings in the field(s) we consider.

There is a long-standing debate in perception science on what principle governs formation of (configural) percepts (Chater, 1996; J. Feldman, 2009; van der Helm, 2000). On the one side, the 'simplicity principle', with roots in the Gestalt tradition, holds that the organization that is as simple as possible given the inputs, will be formed by the perceptual system. The 'likelihood principle', on the other hand, states that the most probable organization consistent with input will be formed. Both positions have important flaws. The likelihood approach has been taken to imply naive correspondence between percept and reality (following empiricist epistemology), as if we can somehow pick the structure in the world that makes the input most likely. Obviously, we have no independent access to the structure of the world[1]. The simplicity approach, on the other hand, seems to assume an inborn, absolute tendency towards simplicity (an *a priori* as found in rationalist epistemology). In practice, however, the simplest, sparsest representation is the one that captures ('explains') the regularities in the input best. Indeed, Chater (1996) showed that simplicity and explanatory power (the extent to which an hypothesis explains inputs) coincide.

Ultimately, neither correspondence to the world, nor simplicity is likely to be a goal in itself for an evolved, biological perceptual system. What is then the criterion for such a system? It can only be what works, i. e.what affords successful behavior in the world. Specifically, a perceptual system that enables an organism to pick up the right regularities in the environment to support survival (maintenance of homeostasis), will prevail. To accomplish this, the brain only has its internal states (neural activity) and how they unfold to go from. It has to learn the structure of these activity patterns, without an external view on the states of affairs to supervise this learning (Clark, 2013b).

This forces the brain to take a proactive role —to self-supervise its learning. The brain constructs the structure it predicts to be out there and the sensory input can then censor these hypotheses. The sensory input serves as feedback on the brain's best guesses. Hence, even 'passive' viewing implies an active, constructive involvement on the part of the brain. Of course, it can also initiate actual movements and witness how input changes as a consequence. So movements can similarly be defined in terms of expected sensory inputs. Both perception and action are then predictive.

There is plenty of evidence for predictive processing in the brain (e. g., Albright, 2012, see also Figure 1). In fact, one might wonder

1 Note that Helmholtz, although considered to be the father of the likelihood approach, never defended such naive realism (e. g., see his 'sign theory' in Helmholtz, 1878).

how else it could accomplish efficient, adaptive behavior. However, the concrete computational scheme the brain uses is still unknown. One contender that is gaining support is called predictive coding. This will be at the heart of the current work, so each theoretical chapter (2, 3, and 6) contains an overview of the predictive coding model, tailored to the specific needs of the issues at stake in that chapter. Predictive coding embraces the (subpersonal) intrinsic predictive activity of every region in the brain, such that a higher-level region predicts and hence explains away activity in a lower-level region. The mismatch or prediction error, understood as the input activity that was not predicted, is used to revise (learn) future predictions.

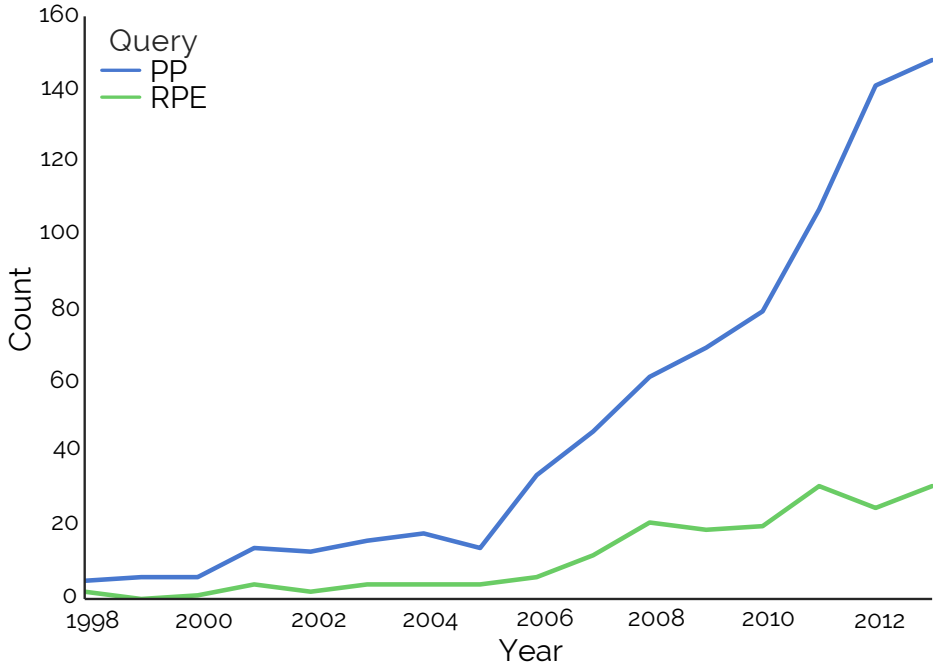

Figure 1: Publications on predictive processing, including predictive coding (PP, blue), in psychology and neuroscience as indexed by the Pubmed database per year. Only part of the increase can be attributed to the success of the related concept of reward prediction error (RPE, green). Landmark publications appeared in 1997 (for RPE, Schultz et al., 1997) and in 1999 (for PP, Rao & Ballard, 1999)

Without going through all mathematical details, we can describe the model, as specified by Rao and Ballard (1997, 1999), as follows. Predictive coding reconstructs a measurable, observable input signal $\mathbf{I}$ (e. g., an image) based on a linear model of a generative matrix

U (e. g., synaptic feedback weights) and an internal (hidden) state vector $\mathbf{r}$ (e. g., neural activity):

$$\mathbf{I} = f(U\mathbf{r})) + \mathbf{n} \tag{1}$$

where $\mathbf{n}$ is a Gaussian stochastic noise process. The $f(x)$ is the nonlinear neural activation function, typically a sigmoidal. The $\mathbf{r}$ vector itself can in turn be predicted by higher level internal states, rendering the model hierarchical:

$$\mathbf{r} = \mathbf{r}^{td} + \mathbf{n}^{td} \tag{2}$$

where $\mathbf{r}^{td} = f(U^h\mathbf{r}^h)$ is the 'top-down' prediction of $\mathbf{r}$, and $\mathbf{n}^{td}$ is again a stochastic noise process. U^h and $\mathbf{r}^h$ are, respectively, the generative weight matrix and internal state vector on the higher level. The latter can represent more abstract stimulus features than the lower level $\mathbf{r}$. For example, in visual perception higher levels can predict (estimate) signal properties at a larger scale by combining responses of a set of lower level units (i. e., they have large spatial and/or temporal receptive fields).

Apart from generating the input image using internal states, we could extend the model to account for how the internal state itself changes with time t. This transition can be specified by a set of recurrent weights V that transform current state vector $\mathbf{r}(t)$ at time t into $\mathbf{r}(t+1)$ at time $t+1$:

$$\mathbf{r}(t+1) = f(V\mathbf{r}(t)) + \mathbf{m}(t) \tag{3}$$

Where V is the state transition (or prediction) weight matrix and $\mathbf{m}$ is a stochastic noise process. This would make it a dynamic model that can represent time-varying input signals (Rao & Ballard, 1997). For simplicity, we leave this extension out in what follows.

The key then is to estimate and learn U and $\mathbf{r}$ from input data only. Estimating $\mathbf{r}$ corresponds to perceptual inference, while learning U entails adapting synaptic weights (perceptual learning). This estimation is possible if we specify an optimization function. To illustrate, suppose we already know the values of U and want to find an optimal estimate for $\mathbf{r}$. Suppose in addition that we already have a prediction $\mathbf{r}^{td}$ for the current state r based on prior inputs. Then the optimization function can be specified as:

$$E_1 = \frac{1}{\sigma^2}(\mathbf{I} - f(U\mathbf{r}))^T(\mathbf{I} - f(U\mathbf{r})) + \frac{1}{\sigma_{td}^2}(\mathbf{r} - \mathbf{r}^{td})^T(\mathbf{r} - \mathbf{r}^{td}) \tag{4}$$

This is the sum of squared errors of level 0 and level 1 (see Figure 2), each being weighted by their respective inverse variances. The larger the noise variance, the smaller the weight of this source of information. Both terms are prediction errors, but the first can in this case be described as the input reconstruction error ($\mathbf{I} - \mathbf{I}^{td}$), while the second is a state prediction error ($\mathbf{r} - \mathbf{r}^{td}$). Hence, the value for the internal state $\mathbf{r}$ that minimizes this function produces the smallest difference between the reconstructed input $U\mathbf{r}$ and the actual input signal $\mathbf{I}$ and deviates as little as possible from the prediction $\mathbf{r}^{td}$ based on prior data.

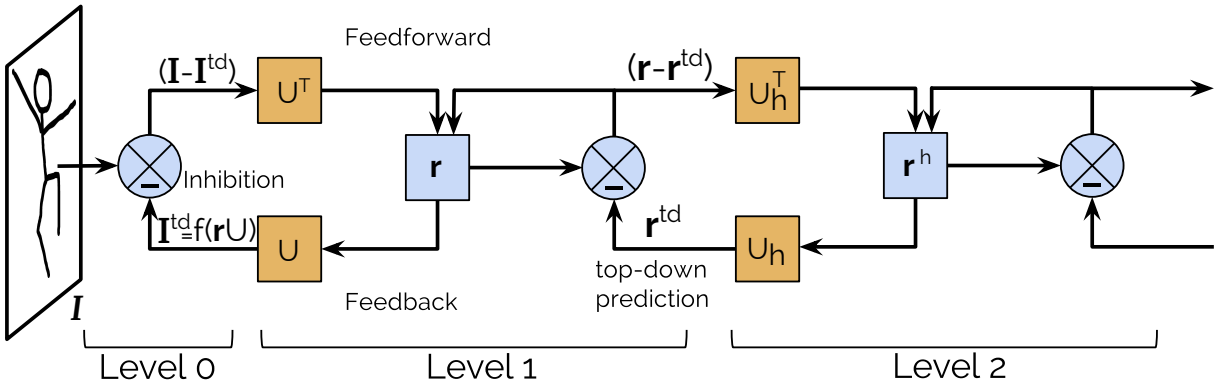

Figure 2: Predictive coding model from Rao and Ballard (1999). Each level consists of several kinds of neurons: feedforward neurons conveying the reconstruction (or prediction) error and encoding the synaptic weights U^T, neurons whose responses $\mathbf{r}$ maintain the current estimate (prediction) of the input signal, feedback neurons encoding weights U and conveying the prediction $f(U\mathbf{r})$ to the lower level, and error-detecting neurons computing the difference $(\mathbf{r} - \mathbf{r}^{td})$ between the current estimate $\mathbf{r}$ and its top-down prediction $\mathbf{r}^{td}$ from a higher level.

Note that this function is just the least squares criterion and it can be minimized with respect to $\mathbf{r}$ or U with gradient descent methods (for full specifications, see Rao, 1999; Rao & Ballard, 1997, 1999). Furthermore, it can be shown (Rao & Ballard, 1997) that minimizing E (the errors) is exactly the same as finding the minimum description length representation of the input signal (assuming Gaussian noise):

$$E = (-\log P(\mathbf{I}|\mathbf{r})) + (-\log P(\mathbf{r})) \tag{5}$$

Hence, E can also be seen as representing the cost of coding the errors and parameters in bits. This means that predictive coding still meets a criterion of simplicity. Note that the right hand side is

exactly the negative logarithm of Bayes' theorem for computing the posterior probability of the state given the input signal (save for the normalizing constant $P(\mathbf{I})$):

$$P(\mathbf{r}|\mathbf{I}) = P(\mathbf{I}|\mathbf{r})P(\mathbf{r})/P(\mathbf{I}) \tag{6}$$

Hence, minimizing E is equivalent to maximizing the posterior probability as described in (hierarchical) Bayesian accounts of perception (Kersten, Mamassian, & Yuille, 2004; Lee & Mumford, 2003). This shows that predictive coding also conforms to the likelihood principle. As in Bayesian accounts of perception, predictions are based on generative, probabilistic models (predictions or beliefs), which are updated based on incoming evidence (prediction errors).

Figure 2 gives a schematic overview of the hierarchical predictive coding model we formally described above. The different levels can be reiterated upwards and are assumed to map onto the cortical perceptual (e. g., visual) hierarchy. The scheme highlights a few key characteristics of the model. First, each level has two populations of neural units: prediction units that represent the current estimate of the internal state, and error units that represent the error signal. Second, feedforward pathways convey the error signals, while feedback connections carry the predictions. Third, top-down predictions inhibit error units one level down, although the model allows for the possibility that predictive inhibitory feedback also stems from local recurrent connections. There is a large ongoing discussion on the biological plausibility of each of these characteristics, but we will largely take them for granted in the current work (although we will briefly refer to the neuroscience linked to predictive coding, see Chapter 2).

In terms of epistemology, then, predictive coding is a clear descendant of pragmatism and its more recent ally, radical constructivism. For pragmatism, to know something is to envision its practical implications and expected sensations (James, 2000[1907]). The knower 'negotiates' a provisional state of affairs with the world, and can only do so by actively constructing its organization. While constructivism (von Glasersfeld, 1991) does not deny the existence the reality of an external world, it just stresses that this Ding-an-sich is off-limits for the knower. Still, we can build up knowledge about it without reference to anything outside of the inputs we receive from the senses. Only through our constructed 'schemata' can we extract new information from the world. The criterion of a correspondence with the world is replaced by 'fit' or coherence (Piaget, 1937; von Glasersfeld, 1995). In predictive coding terms: the fit between the predictions we generate and the activity caused by the

world in our senses. Helmholtz (1878) concurs when writing: "Each movement we make by which we alter the appearance of objects should be thought of as an experiment designed to test whether we have understood correctly the invariant relations of the phenomena before us...". Indeed, we could not know the world if not for these statistical regularities it generates in our sensory epithelia. Predictions are the very fabric of the mind, but, as von Glasersfeld (1995, p. 156) beautifully expresses, "organisms may meet reality only in their failures".

While the move toward active prediction in perception is mandated by the inescapable 'skull-bound' perspective of the brain (the view from inside the black box), it also urges a reconsideration of the perceptual system in the context of a complete biological organism (connecting to the Gibsonian theory of perception; J. J. Gibson, 1979). Specifically, it brings up the issue of how to fit in behavioral relevance and affective value in a theory of perception. As Bialek, Nemenman, and Tishby (2001, p. 2410) remark:

> "... we gather sensory information not for its own sake but in the hope that this information will guide our actions (including our verbal actions). But acting takes time, and sense data can guide us only to the extent that those data inform us about the state of the world at the time of our actions, so the only components of the incoming data that have a chance of being useful are those that are predictive... isolating the predictive information must go a long way toward separating out those features of the sensory world that are relevant for behavior."

As we will see in Chapter 6, even 'simple' single-celled organisms made the move towards prediction for these reasons: mere reaction is virtually always less efficient than predictive preparation. Some of this knowledge (predictions) of what works has been installed by evolution, organisms that did not tap into useful regularities in the environment were just eliminated (Rosen, 2012[1985]). Some of it is learned: new cues that turn out to be predictive of biological concerns. Crucially, the implication for perception is that its goal becomes prediction error reduction, not just *any* uncertainty reduction. It is always the uncertainty relative to the predictions that needs to be reduced. This will form the basis of Chapter 6.

With regard to phenomenal experience, the predictive coding view entails that we perceive what has provided us with successful predictions (including behavior, as perceptual prediction) in the past, given some input cues and context. This could be a rough answer

to Koffka's (1935) central question of why things look as they do. Experimentally, Chopin and Mamassian (2010, 2011) have shown that short-term behavioral success ('usefulness') of a perceptual organization can indeed bias the percept towards this organization. Usefulness was manipulated as task relevance: in a bistable random-dot kinematogram solving a particular task was facilitated when seeing one surface in front of the other. Afterwards, the psychometric function turned out to be shifted in the direction of the interpretation that was more useful before. Based on the simplicity or the likelihood principle, no perceptual bias should be apparent in these bistable geometric figures. Given the dot pattern inputs, both interpretations had the same simplicity and likelihood.

Humans have a lot of flexibility both in the kind of predictions we can build but also concerning when to activate which predictions (Marti, Thibault, & Dehaene, 2014). One could say that we carry around different templates of regularity with which to compare incoming stimuli. Different deviations (prediction errors) will be highlighted depending on the predictions we activate. For example we see a can of soft drink as an object separated from the table on which it stands, because it allows us to pick it up and drink from it. But when we want to know whether it contains more sugar than allowed in our diet, we segregate text columns and words in the label. The rest of the can becomes part of the broader scene. Without much argument, Bateson (1979, p. 49) states that "The division of the perceived universe into parts and wholes is convenient and may be necessary, but no necessity determines how it shall be done." If relevance is a context-dependent measure, we would have to agree with him. This relates to the difficulties defining what an object really is (J. Feldman, 2003).

CHAPTER 2 The way this flexibility in the activation of perceptual configurations is implemented in predictive coding is through what is called *precision*. This will be the subject of Chapter 2. The chapter will give us the chance to examine predictive coding in detail, and to hypothesize what may happen if a key element of this model is absent or inadequately developed. It shows the potential of predictive coding in providing a new account for a clinical disorder, autism in this case, that can explain the subtle pattern of reported evidence in a variety of afflicted mental functions. This is arguably only possible because predictive coding has validity beyond perception, where it was first formulated.

CHAPTER 3 Although we briefly deal with the affective implications of the proposed aberrant predictive coding in autism in Chap-

ter 2, the following chapters will go in-depth on this relation between predictive coding and affect. We start with the specific example of visual art and the emotional experiences art works often elicit. Using well-known art works we look at how predictions are established by the visual cues sculpted by the artist, how they are violated, and how a viewer often manages to return to a state of low prediction errors. We hypothesize that these transitions have affective correlates that make up a substantial part of the aesthetic experience.

CHAPTER 4 This chapter presents an initial effort to experimentally test the hypothesis developed in the context of art. It does not use art works but a stimulus that allows for a more controlled transition in predictability. This will spur a discussion on the links between parsimony of resources (simplicity), Gestalt ideas, and predictive coding.

CHAPTER 5 In this chapter we shift gears and ask whether individuals might activate different (possibly acquired) priors or predictions that can bias their perception, depending on the socio-emotional relevance of those predictions for them. This question directly relates to issues in the literature on emotional biases in perception and attention, but, more importantly for our purposes, it also links to the above mentioned pragmatic view of perception. Similar to the studies by Chopin and Mamassian (2010, 2011) we used a bistable stimulus, of which the two possible percepts were equally likely and simple. We did not manipulate behavioral (or more specifically affective) relevance of the two percepts, but used two groups of participants for which we assumed different relevance of the percepts based on personality characteristics. Hence we expected different perceptual biases.

CHAPTER 6 Finally, we take a broader perspective and discuss how affective value or experience can emerge from predictive coding dynamics. This will require a thorough analysis of the foundations of predictive coding in biology and a broad discussion of how all of this tallies with existing theories on emotion and motivation. We hope this work as a whole shows that it seems be possible, indeed necessary, to give affective value a place within the predictive coding framework if it is to become a unified theory of the mind-brain, but that this may defy some of our intuitions about what emotions are.

[Funes] was disturbed by the fact that a dog at three-fourteen (seen in profile) should have the same name as the dog at three-fifteen (seen from the front). His own face in the mirror, his own hands, surprised him on every occasion ... He was the solitary and lucid spectator of a multiform world which was instantaneously and almost intolerably exact ... he was not very capable of thought. To think is to forget a difference, to generalize, to abstract. In the overly replete world of Funes there were nothing but details, almost contiguous details.

— Jorge Luis Borges, 1942

2

Precise minds in uncertain worlds

Autism Spectrum Disorders (ASD) refer to a group of neurodevelopmental conditions with an early onset, and characterized by socio-communicative impairments and stereotyped, restricted behavior patterns and interests (DSM-5, American Psychiatric Association, 2013). Although ASD has a strong polygenetic component with heritability around 70% (Geschwind, 2011), no biological marker is available yet and thus, diagnosis mainly relies on behavioral assessment. The prevalence of ASD is estimated to be 1%, with males being more affected than females (Baird et al., 2006; Pinborough-Zimmerman et al., 2012). ASD is associated with increased comorbidity for other disorders (e. g., ADHD, anxiety disorders, tic disorders, learning disabilities and epilepsy) (J. L. Matson & Nebel-Schwalm, 2007). In addition, a significant proportion of the ASD population is intellectually disabled (Elsabbagh et al., 2012).

The neurocognitive frameworks put forward to account for behavioral symptoms in ASD can be broken down into two groups, depending on which symptoms they consider to be central and preceding the others. Social first theories put problems with social cognition or motivation front and center. The most prominent contender is the Theory of Mind framework (Baron-Cohen et al., 2000) focus on the social problems and argue that the core deficit lies in the understanding of the behavior of others in terms of their underlying mental states. Nonsocial theories, on the other hand, consider general cognitive or perceptual problems to be the primary causal factor. Among them, the Weak Central Coherence (WCC) theory (Frith & Happé, 1994; Happé & Frith, 2006; Happé & Booth, 2008) and the Enhanced Perceptual functioning (EPF) theory (Mottron & Burack, 2001; Mottron, Dawson, Soulières, Hubert, & Burack, 2006)

focused on the perceptual peculiarities in ASD and argued for a locally (as opposed to globally) oriented processing style in individuals with ASD. Accounts which, prompted by the symptom cluster of repetitive and inflexible behavior patterns, situate the core deficit in an executive dysfunction (e. g., E. L. Hill, 2004), also belong in this group of nonsocial theories.

These theoretical frameworks are not mutually exclusive but focus on different behavioral symptoms. Each theory was highly influential in shaping the field, shifting the research and clinical focus from an exclusively descriptive behavioral approach towards an enhanced desire to understand the atypical neurocognitive mechanisms in ASD. Nevertheless, serious limitations with these frameworks have become evident over the years. First, whilst local processing styles, Theory of Mind difficulties and executive problems are common in ASD, they are neither specific to the disorder, nor are they universally apparent in all cases. Second, although they are called neurocognitive, they do not readily connect to underlying neural mechanisms, except in terms of broad networks of neural activation associated with each domain of function. Part of the problem is a lack of specificity in the proposed cognitive mechanism. Finally, while each of these frameworks attempts to incorporate more than the symptom cluster or behavior on which it is based, this often seems contrived, precisely because each theory is too closely intertwined with the cluster of symptoms in question.

We will argue that the way in which individuals with ASD process and respond to errors (or violations to their predictions) provides an excellent candidate for a primary dysfunction which, when viewed in the context of a complex developmental trajectory, provides a mechanistic explanation for the different symptoms of ASD. This imbalance in the brain's handling of prediction errors could result from different genetic and neurophysiological pathways, thus highlighting that different pathogenetic factors could in fact contribute to a common information processing imbalance (Geschwind, 2011).

We structured the current article as follows. In the first section (Section 2.1) we briefly introduce the predictive coding framework as it originated from perception research, but evolved into a unifying theory of brain functioning. In the second part (Section 2.2) we propose a specific etiological mechanism for ASD, which will then be applied to the different symptom clusters and clinical observations in ASD. Because of the developmental nature of the disorder, we start with a discussion of exploration and development (Section 2.3). Next, we discuss how perceptual and cognitive alterations in ASD can originate from our theory (Section 2.4). In the subsequent sections, sensorimotor and affective consequences will

be covered (Section 2.5 and Section 2.6). In Section 2.7 core principles from earlier sections come together to explain problems in social functioning in ASD. Then, we briefly consider possible neural substrates of the proposed cognitive deficit (Section 2.8). Before reaching our conclusions, we cover a few related approaches of ASD to discuss commonalities and indicate the added value of our approach (Section 2.9).

2.1 THE ANTICIPATING BRAIN

Prediction is central for adaptive, intelligent systems (Hawkins & Blakeslee, 2004). It allows us to efficiently prepare for impinging circumstances that may foster or threaten continued subsistence. However, prediction-based computations can only succeed when there are in fact reasonably predictable contingencies in the world. Prediction, therefore, depends upon an animal's sensitivity to statistical regularities in the environment and in its interaction with it. Some of this structure is readily available, other parts are accessible only through higher order correlations. Our understanding of the role of predictions in shaping information processing has recently taken a step forward by the development of 'predictive coding' models (Clark, 2013b). This computational scheme is heavily inspired by perception-as-inference (Helmholtz, 1962 [1910]) or perception-as-hypothesis (Gregory, 1980) ideas, which assume that the brain continually generates predictions on what input comes next based on current input and learned associations. Predictive coding, however, does not just stipulate that predictions are generated, but that these predictions are compared (at many levels of the system) to incoming sensory input, and that the comparison leads to the computation and representation of an error signal. These prediction errors are important, because they signal that the current generative model of the world —the one used to generate the currently best prediction— is not up to the task of explaining (predicting) the world. Once a prediction error has been signaled, the system still has to employ some degree of flexibility in deciding what do to with that error signal. In an uncertain world, experienced via an inherently noisy biological processing system, errors will sometimes be spurious and uninformative. Thus, whilst prediction errors should sometimes be taken very seriously in updating one's predictive model, it is also critical that some prediction errors are essentially ignored. It is in the imbalance between these options that we think the symptoms of ASD are to find their cause.

In terms of neural architecture, predictive coding assumes a dual computational role for every level of processing (Egner, Monti, & Summerfield, 2010). Representation units compute predictions that

are fed back, while prediction error units compute the difference between sensory input and top-down prediction. These prediction errors then serve as feed-forward input for the next level. The biological plausibility of this specific architecture is still under investigation, but the importance of prediction errors and predictive processing in the brain in general is well-established. Predictive coding can account for fundamental stages of perceptual processing, such as the emergence of extra-classical receptive field effects measured with single cell recordings in the primary visual cortex (Rao & Ballard, 1999). It can also account for the complex dynamics between predictions made and input received at very different stages of the system (den Ouden, Daunizeau, Roiser, Friston, & Stephan, 2010). Furthermore, it can explain neural dynamics such as the apparent adaptation to predictable stimulus contingencies (Summerfield, Trittschuh, Monti, Mesulam, & Egner, 2008). Finally, there is some evidence for the existence of separate representations for input and error signals in the recent discovery of differential sensitivity to predictable stimuli in separate clusters of voxels in the Fusiform Face Area (de Gardelle, Waszczuk, Egner, & Summerfield, 2012).

The computational scheme of predictive coding is assumed to repeat on every level of the perceptual hierarchy (Diuk, Tsai, Wallis, Botvinick, & Niv, 2013; Wacongne et al., 2011). Each higher level can capture a higher order regularity in input, relating events spanning more time or space, because it can work on the representational 'language' of the previous level. Perceptual inference is guided in a top-down way through higher-level, conceptual predictions that can be passed downwards generating a chain of interdependent predictions to match on different levels, from complex features to low-level stimulus characteristics.

Formally, predictive coding is equivalent to Bayesian inference with the priors replaced by predictions and sensory evidence replaced by prediction errors, reflecting the mismatch between input and predictions. However, the differences between these two related approaches have important implications. A first distinction from Bayesian approaches concerns the more specific claims about the neural implementation of predictive coding. Second, replacing sensory evidence by prediction errors emphasizes that incoming information is put in context from the very start. It immediately becomes input relative to the organism, its models of the world and its current state. It also emphasizes that processing does not start with the onset of a stimulus. Pre-existing, intrinsic activity of the brain is considered formative as it reflects the continuous predictive activity of the pro-active mind-brain (Bar, 2009). Another advan-

tage of predictive coding is that it allows a natural connection to other neurobehavioral domains, where prediction errors are known to play a crucial role, like midbrain dopaminergic processing of reward (Schultz et al., 1997), hippocampal processing for contextual memory (Honey, Watt, & Good, 1998) and amygdalar processing for fear learning (Boll, Gamer, Gluth, Finsterbusch, & Büchel, 2013). This suggests we may be a step closer to a general theory of the brain as a prediction engine in which prediction errors emerge as the lingua franca of neural information processing (den Ouden, Kok, & Lange, 2012).

Critically to our theory of ASD, predictive coding operates on two time scales (Dayan, 2012b; Friston, 2010a). Predictions are used here-and-now to shape one's online estimation of the state of the world (albeit through an iterative process), but the resulting prediction errors also shape plasticity and learning over longer time scales. In this way, today's prediction errors shape tomorrow's predictions (paraphrasing a famous Bayesian dictum). Because the world is not static, predictable contingencies that used to hold can change, and predictive coding has to track these dynamics. No two experiences are ever completely the same, thus, prediction error will always be present to some degree. However, the brain has no direct, independent means of differentiating mere noise from actual changes in the world (J. Feldman, 2013). It is, therefore, critical that predictive coding incorporates a mechanism to flexibly alter the extent to which the prediction errors generated by online estimation affect future learning and plasticity.

A solution to this can be found in terms of a flexible adjustment of what Friston (2010a) describes as the *precision* of the prediction errors. To explain precision, one can draw the parallel with the means comparison in a t-test, in which the numerator represents the prediction error, which is weighted by the estimated standard error (precision or confidence) (Friston, 2009). As in the t-test, precision is not given in perceptual inference, it has to be estimated as well. In an optimal system, precision has to increase when there still are learnable regularities in the environment, and decrease when it is estimated that remaining deviations can be attributed to noise that is unlikely to repeat in next instances or to other irreducible uncertainties in input. Distinguishing between irreducible and reducible uncertainties is a fallible process, relying on complex meta-predictions for a given context. The system, therefore, has to attribute a value or weight to prediction errors in order to determine to what extent they should induce new learning. The role of precision is conceptually the same as that of the learning rate parameter in Rescorla-Wagner learning (see Courville, Daw, & Touretzky, 2006; O'Reilly, 2013, for a full

discussion on learning in volatile environments). Setting precision consequently relies on a form of meta-learning: learning what is learnable (Gottlieb, 2012) or estimating the predictability of new contingencies. It is clear from all of this that precision should be a context-sensitive measure, to be flexibly optimized dependent on the current class of input and the state of an organism. Indeed, precision is assumed to be the mechanism of attention within predictive coding. At its core, attention is the process of deciding where to look next, to allocate resources to that information with the highest value, understood precisely as input containing reducible uncertainty (Dayan, Kakade, & Montague, 2000; Gottlieb, 2012). Neurally, precision is assumed to be represented by the gain of bottom-up neural units representing the prediction errors, probably mediated by neuromodulators (Friston, 2009; see Section 2.8).

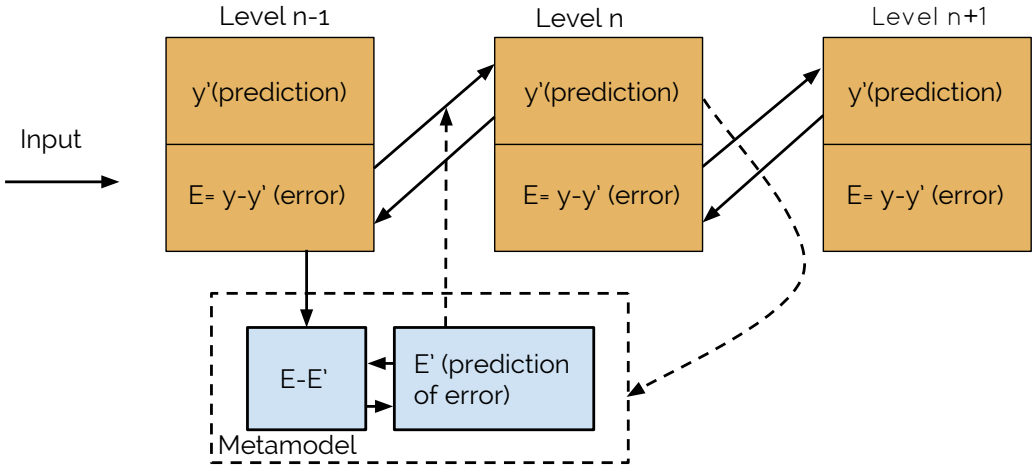

Figure 3: Sketch of the predictive coding scheme. To avoid clutter, we only depicted a metamodel for one predictive level. Note that the prediction of error is dependent on the prediction that generates the error (curved dashed arrow). This prediction of error (the precision) modulates the impact (gain) of the prediction error on the prediction units (straight dashed arrow).

From this brief overview it should be apparent that predictive coding provides a framework that allows us to go beyond unidirectional views of information processing. Bottom-up information streams (predictions errors) are inherently dependent on top-down influences (predictions), which in their turn are shaped by previous prediction errors. This complex interplay also means that the dysfunction of one will automatically have consequences for the other. Disturbances in the relative contribution of top-down versus bottom-up information flow have been at the heart of two influential

cognitive theories of ASD, representing apparently diametrically opposing positions (WCC and EPF). A predictive coding approach provides a principled and refined view on the influence of top-down versus bottom-up processes and their complex interplay.

2.2 PREDICTIVE CODING IN ASD

To bring into focus what we believe is the core processing deficit in ASD, we have to emphasize again the distinction between reducible and irreducible uncertainty (prediction errors). Irreducible uncertainty is due to the inherent stochastic nature of the world and the inherently noisy biological apparatus with which we sample from that world. Differentiating between reducible and irreducible uncertainty requires an estimation of 'expected uncertainty' based on previous prediction errors (Preuschoff & Bossaerts, 2007; Yu & Dayan, 2005). If through learning, you estimate the outcomes of a stochastic process to vary with $\pm$ 3 (hypothetical) units, a prediction error of 2 should not surprise you, and therefore should not urge you to update your model (prediction). When the size of a prediction error is smaller than the expected variability (based on past prediction errors) for this event, the current prediction error should be scaled down. Reducible uncertainty, on the other hand, is present when associations in the world (or our interaction with it) are not fully learned yet. The latter is sometimes called unexpected uncertainty. This is the case when previously predictive cues have changed and become invalid, so a real update of the model is necessary. More formally, it is about situations in which correlations between predictions and prediction errors have changed.

In relatively unambiguous situations, people with ASD can successfully learn and apply new contingencies (M. Dawson, Mottron, & Gernsbacher, 2008). Problems arise, however, when the predictive value of learning cues changes (i.e., in volatile environments). For that reason, we situate impairments in ASD in meta-learning: learning which cues of all present stimuli are learnable, i.e., can reliably predict future situations relevant for the task at hand. This meta-capacity, estimating for which cues predictive progress can be made, allows Typically Developing (TD) individuals to distinguish random variability in input from actual, learnable changes in environmental regularities. Here, we advocate that individuals with ASD overestimate the amount of changes in environmental regularities, because they give too much weight to their prediction errors.

Another way to conceive of this meta-learning capacity is in terms of knowing where gains can be made in predicting the world. If you know where predictive progress can be made, you know which

prediction errors matter, hence which prediction errors should be assigned high precision. Precision is the mechanism of attention in predictive coding because in this way it affects the further sampling of the sensory world. Atypical attention happens to be among the earliest signs of ASD, described in terms of the flexible and appropriate assignment of salience to stimuli (Elison et al., 2013; Zwaigenbaum et al., 2005). In ASD, the atypical distribution of attention has been attributed to slower encoding, which is consistent with the thesis that too many resources are invested in sensory processing because precise prediction errors cannot be discounted and thus attract further processing.

Hence, derived from a general theory of information processing, predictive coding, and our analysis of what could be the key problems in ASD, we situate the core deficit in the High, Inflexible Precision of Prediction Errors in Autism (HIPPEA). Low-level sensory prediction errors are generally set at a level of precision that is too high and independent of context (Palmer, Paton, Hohwy, & Enticott, 2013; Van de Cruys, de-Wit, Evers, Boets, & Wagemans, 2013). As mentioned before, it is useful to consider the consequences with regard to online inference versus those regarding learning separately. If prediction errors during online inferences get an unduly high precision, these will urge new learning for every new event. The predictions that result from this learning will be shaped by noise that is unlikely to repeat in the future, hence these predictions will almost never be applicable. In neural network learning studies, overfitting takes place when errors for the training set are reduced to an exceedingly low level (Bakouie, Zendehrouh, & Gharibzadeh, 2009). It is a suboptimal form of learning because new data (acquired with each new experience) will generate large errors, meaning that there is little or no generalization. If errors are always deemed important, every new instance will be handled as an 'exception', different from previous experiences. In the long run, however, those affected by this dysfunction, may succumb to a sort of learned helplessness: too much learning with no fruits. This may have an especially demotivating effect on particularly 'noisy' interactions, such as those involved in social situations (see Section 2.7).

With regard to the consequences for online inference and behavior we have to distinguish situations in which an exact match from cue to prediction exists and is functional from situations in which exact matches will rarely happen or are even dysfunctional. In the case of exact matching, it is well-known that people with ASD cope incredibly well (Mottron et al., 2013). They often excel in rigid, exact associations (rote learning). Here, their overfitted predictions serve them perfectly well, precisely because they suffer less from interference

from similar instances. They seem to trade off the ability to generalize with a more accurate memory. Hence, according to HIPPEA, the core processing deficits in ASD become most evident when some disregard for details and some generalization is needed. Generalized inferences are required in situations where exact matches are not present, which is the rule rather than the exception in natural situations, especially those involving social interactions.

In everyday life, multiple cues impinge simultaneously on an individual. At first exposure this may cause sensory overload, because selectivity is lost when the informational (predictive) value of cues cannot be estimated immediately. Predictions are tested but violated because they are based on spurious correlations. Individuals with ASD may cope with perceived repeated changes in contingencies by executing prepotent, impulsive or 'model-free' behaviors, described as repetitive, stereotyped behaviors in the ASD symptomatology (for a discussion on the role of precision in arbitrating between model-free and model-based behavior, see Clark, 2013a; Daw, Niv, & Dayan, 2005). In a second stage, individuals with ASD may 'give up', and select cues just to evade and cope with prediction errors. On their own scale, these cues may be highly predictable, even though they are not functional in the situation at hand. Thus, attention and behavior become dominated by one or a few cues (cf. stimulus overselectivity; Lovaas, Koegel, & Schreibman, 1979), singled out seemingly arbitrarily.

Note that computing prediction errors as such is not impaired in ASD according to this view. Individuals with ASD can still compare their predictions with actual input. These prediction errors, however, have to be weighed in accordance to an estimation of their reliability, i.e., the extent to which they are caused by learnable (changes in) regularities. Attesting to the fact that prediction error computations are intact in people with ASD, their detection and discrimination performance seems to be similar to typically developing individuals, if not superior (see Section 2.4.2).

Importantly, one can distinguish between two mechanisms that both can result in inflexibly high precision of prediction errors. First, it is possible that the neural mechanism for precision is directly affected in ASD, fixing precision at a high level and preventing meta-learning (which may take place anyway) to have an effect on perception and learning. Aberrant neuromodulatory mechanisms of precision, as discussed in Section 2.8, may be responsible here. Second, the meta-learning prior to the setting of precision may be deficient in ASD and hence does not provide the needed basis for proper, context-dependent estimation of precision. Neural regions

and mechanisms that may be central for this capacity are discussed in Section 2.8.

In a nutshell, HIPPEA consists of the following basic premises. The starting point is a high *and* inflexible estimation of precision of prediction errors in ASD. This indicates meta-learning is deficient or short-circuited. Indiscriminately high precision will mean that unrepeated, accidental variations in the input receive disproportionate weight. This in turn, prevents abstract representations to be formed, because matching will continue on a more specific level, closer to the input. Indiscriminately high precision also induces superfluous learning, leading to narrowly defined, lower-level predictions, and incomplete hierarchical models. Finally, indiscriminately high precision entails a loss of autonomous, flexible attentional selection based on informativeness (deciding what information to sample based on the different types of uncertainty in input).

HIPPEA thus situates problems in ASD at the intersection of perception, attention, learning, and executive functioning (adaptive control). Further key symptoms of ASD may emerge from this impairment, but this will be fleshed out in the sections below. We argue for an impairment in general information processing rather than in one single domain (e. g., social cognition), supported by the fact that problems in ASD are not limited to one such domain, but are pervasive. However, this also puts the burden of explanation with us as to why some domains (specifically the social) would be affected more than others (see Section 2.7).

2.3 DEVELOPMENT AND EXPLORATION

The meta-learning deficit in HIPPEA is very consistent with the developmental nature of the disorder. The very process of moving from one 'simpler' developmental stage to the next 'complexer' one is impaired when an organism cannot estimate where predictive progress can be made. If any prediction error is deemed as valuable as the next, an inappropriate lingering on stimuli is expected to occur. As a result, the kind of exploration that optimizes learning is lacking, because estimating where predictive progress can be made helps an organism to avoid the large regions of input space that cannot be learned (fully) and those that are too difficult at this stage of development. In short, this principle gives a rationale for the importance of intermediate levels of complexity in development (Berlyne, 1966; E. Gibson, 1969; Oudeyer, Baranes, & Kaplan, 2010). If predictive gain can be properly estimated, exploration can be guided such that it is aimed at regions with a difficulty just above current ability, which leads to discernable progressive stages of increasing

complexity, as modeled in developmental robotics (Oudeyer, Kaplan, & Hafner, 2007). Particularly in noisy, variable environments the mechanism can be expected to realize more efficient learning. It is easy to see that if this capacity for active exploration is missing, as we think is the case in ASD, an individual has to rely much more on the scaffolding provided by caregivers, explicitly guiding progression from simple to more naturalistic situations. Apart from prenatal genetic and neural components, differing degrees of this environmental scaffolding may account for heterogeneity in symptom severity and developmental trajectories in ASD.

The link between prediction violation and exploration is elegantly illustrated in a study by Legare (2012), investigating how TD children explain evidence violating their predictions and illustrates how this mechanism may shape development. Different shapes were put on top of boxes that could light up, depending on the shape, and those shapes that caused the box to light up were subsequently labeled as a "blicket". Children were then confronted with a violation of the established prediction (no light for a blicket) and Legare (2012) asked them to explain what had happened. She could distinguish two main types of explanations; about half of the children tried to explain *why* the block did not light up (e. g., no batteries, block was not placed properly), while a third of the children explained the situation by referring to the category membership (e. g., "It's not really a blicket; it only looks like one"). Most interestingly, however, the kind of explanation children gave predicted the way they played with the objects later on. While children who gave a causal explanation explored the objects more thoroughly, testing different combinations and experimenting with the placement and orientation of objects to find out what would happen, children who explained inconsistency in terms of the categories primarily went about sorting the objects in two different categories based on what had happened when they first placed them on the box. This sorting behavior was a less sophisticated form of exploration, and less likely to foster deep understanding of the underlying sources of inconsistencies.

Arguably, the difference hinges on the ability to model uncertainty in associations in the input. This modeled uncertainty becomes a handle to dissect underlying causes. The precision of low-level inconsistency can, with a proper model of uncertainty, be down-regulated such that the general rule (prediction) is not violated and so does not have to be abandoned. Rather, modeling uncertainties in the task opens the door to contextual modulations of the general rule, which are always at play in practice. When uncertainty is not accounted for, and precision is continuously high, every mi-

nor violation will induce new learning. An inconsistent finding is categorized anew or considered a 'special case' unlike previous instances. The latter is what HIPPEA proposes to be the case in ASD. Though Legare's (2012) study only included TD children, the sorting behavior found in spontaneous play for the subset of children that gave non-causal explanations, is reminiscent of what is observed in autistic play. Her results show that whether and how people explain prediction error is linked to the kind of exploration they will engage in. In our line of thinking, the difference already emerges in the way people process perceptual input that runs counter their predictions and this may have far-reaching consequences for exploration and further development, notably with regard to finding out about why the world functions as it does.

Considering this change to the nature of exploration in ASD, it is informative to revisit the so-called 'dark room problem' within predictive coding (Friston, Thornton, & Clark, 2012; Froese & Ikegami, 2013). This problem arises because if, as the fundamental thesis of predictive coding has it, an organism acts to minimize the prediction errors it experiences, then the simplest solution would be to seek out a dark room, devoid of prediction errors. Nevertheless, most organisms venture out of their black boxes and explore the world. The most obvious way to counter this is by noting that generalized predictive coding not only involves learned mental models and perceptual predictions, but also bodily predictions, predictions embodied by the very structure of the body, homeostasis, biomechanics and the "gross initial neural architecture of the agent" (Friston, Thornton, & Clark, 2012). Evolution equips organisms with a limited set of expected states (cf. homeostasis) that is compatible with their continued existence (survival). A dark room will not remain a low prediction error environment, for instance, when food is not available.

One means of ensuring that organisms venture out to meet their needs for survival, is to equip them with an ability to tolerate the prediction errors associated with new unpredictable environments. Organisms can adjust the precision of prediction errors based on the expected volatility of their environment. If precision of low-level prediction errors is overly high, however, the organism may very well prefer to stay in a dark room-like environment. In fact, the typical autistic state of stereotypic self-stimulation and indifferent withdrawal from the world and from others, can be regarded as "abnormal yet effective ways of reducing prediction errors" (Froese & Ikegami, 2013, p. 213). By their caregivers, children with ASD are often decribed as detached from the world, as if they are living in their own walled world. Not because they are unhappy, or unable

to move or sense, but because they are satisfied with the current level of complexity of the environment. The prediction error minimization principle says that "we harvest sensory signals that we can predict" (Friston, Thornton, & Clark, 2012). Hence, it seems that children with ASD, since they (initially) cannot predict more complex environments, are perfectly content to stick to the confined space and motion they know.

2.4 COGNITIVE FUNCTIONING

In the following sections we review the most relevant literature illustrating the implications of prediction errors with chronically high precision in cognitive and perceptual domains. At the end, we will also describe the special perceptual and cognitive skills that some individuals with ASD have developed ('savant skills'), which can result from the potential benefits of high-precision prediction errors when applied to certain domains. The problems in reasoning about mental states ('mentalizing'), which are also a central aspect of cognition in ASD, will be covered in Section 2.7 on social functioning.

2.4.1 *Attention and executive functioning*

An interesting pattern of findings has emerged from attention studies in ASD, comprising both superior performance in certain tasks and severe deficits in others. Below we substantiate that the specific pattern is largely consistent with HIPPEA. We start considering visual search tasks, then move to more complex attention tasks with a larger executive component, and finally make new predictions based on our account and propose suitable designs to test these.

Visual search studies demonstrate that performance on some attentional tasks can be intact or even enhanced in ASD. Superior visual search has been found both when the target is defined by a single feature and for conjunctive targets (Keehn, Müller, & Townsend, 2013; O'Riordan, Plaisted, Driver, & Baron-Cohen, 2001; Plaisted, O'Riordan, & Baron-Cohen, 1998a). Moreover, performance seems to correlate with symptom severity (Joseph, Keehn, Connolly, Wolfe, & Horowitz, 2009). Group differences are especially present in conjunction search tasks or tasks with higher difficulty. A predictive coding account of visual search would start from the predictability within search displays. When every item in a display reinforces a particular 'prediction', an anomaly (the 'odd one out') causes an 'error' that becomes salient. Heightened precision of this prediction error means enhanced salience of this oddball, which facilitates quick

detection. Thus, individuals with ASD seem to be just as good or even better at exploiting predictability in a display.

In more complex attentional settings however, performance usually declines substantially in autism. As we saw, precision (or weight) of prediction errors should be flexibly adapted based on meta-learning (learning what the relevant features in a task are). When precision of prediction errors is uniformly high, the selective force is lost when processing a context with multiple cues. Hence, difficulties in allocating attention may be expected. Phenomenally, this may manifest itself as attention to irrelevant features, on the one hand, and as lack of disengagement or perseverative attention, on the other hand. Yet, this problem occurs only when multiple cues compete. If only one cue is present, ensuring that the selection process is clearly imposed by the task itself, performance remains intact (Burack, 1994).

An experiment by Pierce, Glad, and Schreibman (1997) confirms this. When ASD children, TD children and mentally disabled children were presented with video fragments of social interactions containing one or more cues, ASD children did worse than the other two groups when asked to answer a set of questions right after, but only in the multiple cue conditions. We argue the problem is one of autonomous selection; the relevance or redundancy of the cues is not recognized. Consistent with this, task performance in ASD is expected to suffer most when it is dependent on autonomous exploration and efficiently probing of available cues rather than fixed instruction (clear top-down selection).

Others before us (e. g., Elsabbagh et al., 2009; Keehn et al., 2013) have situated the origin of problems in ASD in attentional difficulties, more specifically in disengaging attention. However, we consider these disengagement problems not as primary but as an effect of the lack of adaptive precision of prediction errors. This kind of overselective (perseverative) attention does not stand in contradiction with what we said before on the lack of autonomous selectivity in ASD. It is the flexible adapting of selectivity in a task-dependent way that is lacking in ASD. Uniformly high precision will create a prolonged processing of all stimuli (and an associated deficit in disengaging). This is also apparent in studies by Sasson and colleagues (Sasson, Elison, Turner-Brown, Dichter, & Bodfish, 2011; Sasson, Turner-Brown, Holtzclaw, Lam, & Bodfish, 2008), demonstrating perseverative attention (longer fixation times per image explored) and more intensive, detail-oriented exploration of a limited number of images in ASD. Hohwy and Palmer (2014) note that increased precision could lead to longer sampling of incoming signal in order to attain the precise signal people with ASD expect, before making a decision. If so, such longer sampling may as well help to explain

larger reaction times for diverse tasks, often reported in ASD. In any case, we surmise that lacking disengagement is not the core mechanism but rather one of the consequences of HIPPEA. However, often perseveration and overselectivity may be strategically replaced by avoidance and underreactivity.

It is clear that the proposed difficulties in autonomous cue selection will cause broader problems in executive functioning, in particular with regard to cognitive flexibility or set shifting. According to the executive functioning theory, the latter functions are impaired in ASD and this impairment is assumed to underlie the restricted, repetitive and stereotyped pattern of behavior and interests (Lopez, Lincoln, Ozonoff, & Lai, 2005). Though problems with flexibility have clearly been found in daily life (Gioia, Isquith, Kenworthy, & Barton, 2002; Mackinlay, Charman, & Karmiloff-Smith, 2006), studies measuring cognitive flexibility in a clinical or research setting have yielded less consistent evidence. Overall, studies using the Wisconsin Card Scoring Task (WCST) report clear deficits, reflected by a higher number of perseverative responses when a rule switch should occur, while more controlled task-switching paradigms generally fail to find cognitive flexibility problems in ASD (Geurts, Corbett, & Solomon, 2009; Van Eylen et al., 2011).

Recently several researchers suggested that these inconsistencies may be due to differences in the extent to which explicit task instructions are given, denoted as open-endedness (Van Eylen et al., 2011; White, 2013). When task instructions contain no explicit indication of the rules to be applied, nor that a rule switch will occur (as in the WCST), results show rather consistent cognitive flexibility deficits in ASD. In this case, participants have to be able to autonomously filter-out and focus on relevant information in a situation where multiple cues compete. There is evidence that individuals with ASD have difficulties doing so and overly focus on irrelevant, often low-level details (Stoet & López, 2011). In contrast, when a cue explicitly indicates which rule to apply and when to switch, all studies report intact performance in ASD. Hence, the act of switching does not seem to be a problem per se (Poljac & Bekkering, 2012).

All this is very compatible with our interpretation of ASD in terms of an overweighing of prediction errors. As we saw, the informativeness of cues has to be derived from meta-models, which should adjust the precision with which errors based on these cues are weighed. A loss of this capacity would lead to a deficit in the ability to autonomously select cues that have predictive value in situations where multiple cues compete. Learning a new unambiguous contingency in itself is not a problem, but individuals with ASD struggle with spontaneously noticing that the predictive value of particular

information changed. This leads to cognitive flexibility deficits on open-ended tasks, but not on tasks were explicit instructions are provided about what is informative and when. Testing a range of executive functions in ASD, White, Burgess, and Hill (2009) corroborated that all open-ended tasks generated group differences, while none of the more constrained tasks did. Hence, this reasoning might also explain some of the inconsistencies in studies of other executive functions (Gioia et al., 2002; White et al., 2009).

Open-ended, generative sorting experiments provide converging evidence. For example, in a free sorting task with children books (Ropar & Peebles, 2007), ASD children relied less on category labels (games versus sports) and more on purely perceptual features (color and size) than TD children. More one-dimensional sorting was found in free sorting of shapes by children with ASD, especially in more complex stimulus sets (D. J. Edwards, Perlman, & Reed, 2012). In a twenty-questions game, children with ASD consistently generated questions (predictions in our context) of lower quality, especially more concrete ones that eliminated fewer items at a time (Alderson-Day & McGonigle-Chalmers, 2011). Analyses indicated that difficulties in managing relevant and irrelevant information were likely sources of the problems of children with ASD. This cognitive control problem, which is at the heart of HIPPEA, also explains why individuals with ASD are particularly slower in early blocks of categorization learning, when flexibly switching the focus of attention from one dimension to another dimension is needed (e. g., Soulieres, Mottron, Giguere, & Larochelle, 2011).

To clearly summarize our hypothesis: When real environmental changes go together with random changes, disentangling the two is particularly difficult for people with ASD. They seem to be able to learn changes in contingencies, when they are clearly indicated, as in some set shifting tasks. Similarly, they can learn fixed contingencies, even in probabilistic environments and without explicit instructions, as implicit learning studies show (J. Brown, Aczel, Jiménez, Kaufman, & Grant, 2010; Nemeth et al., 2010; Pruett et al., 2011). However, these two combined create the clearest deficits. Therefore, we hypothesize that adding noise by using a probabilistic switching task, would increase their flexibility impairments. This has indeed been observed by D'Cruz et al. (2013) in a reversal learning task with intermittent non-reinforcement. Moreover, these switching problems correlated with severity of repetitive and restrictive behaviors. From our perspective, this kind of task will be most sensitive in picking up deficits in executive functioning for ASD.

Although these findings are largely compatible with the prediction derived from HIPPEA, future attentional studies should test our

hypothesis more directly. A modified version of Posner's attention cueing task as developed previously (Vossel et al., 2014; Yu & Dayan, 2005) could contribute to this. In the typical Posner cueing task, a simple cue (a briefly presented flash) only indicates the actual (valid) location of the target in a certain percentage of trials (e. g., 75%). Typically participants will learn to use the cue information to improve their detection speed to the extent that the cue is reliable. This improvement may also be present for individuals with ASD, but we predict that things will go awry in ASD when the probabilistic structure changes unexpectedly during the experiment, for instance, when the predictability of a cue changes across blocks. In such a volatile environment, the validity of the cue (the extent to which it predicts the target location) varies over the course of the experiment. Prediction errors usually lead to the updating of beliefs (predictions) about the environment, but the impact of these prediction errors should be tuned to whether additional learning is expected to be still possible. In a fully learned stable phase, new prediction errors are probabilistic noise that should lead to little or no update of predictions. However, when new learning is estimated to be possible, for example when probability structure changed, recent prediction errors should significantly update current predictions. This task shows the importance of contextual, flexible setting of precision.

Another variation of the Posner task that could provide a useful test of our theory has been developed by Yu and Dayan (2005). In this version, a set of cues (e. g., differently colored arrows pointing left or right) precede the target. For any one trial, one particular cue (color) from the set predicts the target location with a certain probability (e. g., >.5). This cue type and validity remain active for a considerable amount of time, creating a stable environment. Then, unbeknownst to the participant, this context is suddenly changed: A different cue now predicts the target location with a different cue validity. Note the similarity with traditional set switching tasks, although the rules there usually are deterministic, rather than probabilistic. Participants with ASD will have distinct problems with this task, again because two forms of uncertainty are pitted against each other, as described above. An added benefit of these tasks is that a hierarchical Bayesian model can be used to quantify precision (or learning rate) on a trial-by-trial basis (Behrens, Woolrich, Walton, & Rushworth, 2007; Yu & Dayan, 2005), and on a subject-dependent basis (Mathys, Daunizeau, Friston, & Stephan, 2011; Vossel et al., 2014), pin-pointing exactly whether and how the learning style of ASD subjects differs from that of TD individuals.

From the above it should be clear that HIPPEA has a natural way of explaining the discrepancy between the experimental data in contrived laboratory contexts and the clinical observations in daily life. As most ASD researchers know, it is surprisingly difficult to find statistically significant group differences in the lab that should occur according to everyone's expectations based on the major problems that people with ASD experience every day (J. L. Amaral, Collins, Bohache, & Kloos, 2012). Natural circumstances are often much more unpredictable and open-ended with lots of accidental variability, and hence lead to clear deficits in ASD (Kenworthy, Yerys, Anthony, & Wallace, 2008). The lab, in contrast, usually provides a well-controlled environment, in which it is made very clear what is expected (explicit instruction, practice trials) and with multiple instances of the same (often simple) task (repeated trials). Little autonomous control is needed here. Where many TD children easily get bored in such a context and start talking to the experimenter, kids with ASD usually like these repetitive, computerized tasks, and they are motivated to do well in them.

2.4.2 *Perceptual processing*

Research on visual processing in ASD has been dominated by two related theoretical frameworks that each emphasized a different side of the coin: WCC theory emphasized reduced global processing (Frith & Happé, 1994), while EPF theory emphasized enhanced local processing (Mottron & Burack, 2001). More recent accounts describe the peculiar aspects of visual processing in ASD more in terms of a bias or perceptual style, a disinclination for global or a preference for local processing (Happé & Frith, 2006; Happé & Booth, 2008; Mottron et al., 2006). Despite a vast amount of research on visual perception in ASD, the atypical profile of visual processing is only partly understood, and the empirical evidence for the original ideas is mixed (for recent reviews, see Behrmann, Peterson, Moscovitch, & Suzuki, 2006; Dakin & Frith, 2005; Simmons et al., 2009).

HIPPEA is compatible with both EPF and WCC, but it offers a more specific foundation, describes dynamics in learning and inference, and hence has different implications. According to HIPPEA, precision of bottom-up information is uniformly amplified, an idea that is consistent with EPF, but we can better specify how and why perception is enhanced. The detectable size of prediction errors is not smaller, but rather the weights (precision) these errors receive are higher. HIPPEA does not reduce problems to a purely bottom-up way of perceptual processing. Because it is embedded in the inherently bidirectional predictive coding framework, the mutual, constructive

interaction of bottom-up and top-down information flows is central. Specifically, increased precision of prediction errors will have important consequences with regard to the kind of predictions that will be formed based on prediction errors with unusually high precision. Perceptual inference and learning will not progress to higher-level, more abstract representations because of the emphasis given to violations to those higher-level representations at lower levels of processing. Learning will result in predictions tuned sharply to exact perceptual input cues. As a result, primarily low-level predictions will be formed, which will have limited applicability, while higher-level predictions will be triggered less automatically by incoming information, an idea that is consistent with WCC.

In ASD, stimuli are treated in an idiosyncratic manner, because slight deviations are perceived as informative and all experiences are thus more readily treated as new instead of as belonging to a known category. More concretely, the focus on prediction errors at lower levels causes individuals with ASD to focus on concrete but irrelevant changes in viewpoint or illumination, which impede the ability to progress to the more relevant, abstract levels of description in terms of shape or object identity. Note however, that the predictive machinery in ASD is not deficient in our view: Predictions are still formed and prediction error is computed correctly. Hence, global interpretations are not necessarily lost in ASD; they just require more experience and they will appear only under more constrained conditions. So, while a familiar representation may not pop-up automatically when a related stimulus appears, top-down activation of holistic, Gestalt-like templates and global processing are often still possible, but as a conscious strategy, when task instructions require it and enough time is available. For individuals with ASD, it is not the default, automatic processing mode. This accords nicely with the recent move in the field towards differences in default preference or bias (often measured by initial choice responses or reaction times) rather than in distinct inabilities (measured by error rates). This interpretation receives support from a recent meta-analysis of the mixed evidence from a variety of local-global perceptual processing tasks, which demonstrates that global processing takes time in individuals with ASD (Van der Hallen, Evers, Brewaeys, Van den Noortgate, & Wagemans, 2014). Moreover, the inconsistencies in the literature also make sense in this perspective. Laboratory tasks mostly use standardized stimuli, and often do not incorporate the noise that is usually present in real-life stimuli. In these constrained circumstances, individuals with ASD can actually perform on a typical level.

2.4.2.1 *Low-level perception*

According to HIPPEA, low-level differences will get boosted and sent upwards influencing behavior and learning. Setting precision high by default may give an advantage for lower-level processing (but impedes building and using of a hierarchy of predictions for generalization). In the auditory domain this is reflected in the frequently reported enhanced pitch perception in children and in a subgroup of adolescents and adults with ASD, especially those with early developmental language delay and language-related difficulties (for reviews, see Haesen, Boets, & Wagemans, 2011; O'Connor, 2012). Superior pitch processing has been established regardless of stimulus complexity (i.e., pure tones, complex tones, speech sounds, nonwords, words) using a variety of psychophysical tasks (e. g., identification, discrimination, memory) (e. g., Bonnel et al., 2010, 2003; Jones et al., 2009). Relevant in this context is also the increased prevalence of absolute pitch and musical savants in the ASD population (e. g., Heaton, Williams, Cummins, & Happe, 2008).

In visual perception, findings are more mixed. Most studies have found little or no group differences for visual acuity (Simmons et al., 2009). One study observed a small group difference indicating superior contrast sensitivity in individuals with ASD (Bertone, Mottron, Jelenic, & Faubert, 2005). Another study found evidence for superior visual acuity (Ashwin, Ashwin, Rhydderch, Howells, & Baron-Cohen, 2009), but has been disputed on methodological grounds (M. Bach & Dakin, 2009) and replication attempts have failed (Bölte et al., 2012; Keita, Mottron, & Bertone, 2010). Based on HIPPEA, however, there is still potential for well-controlled studies to find detection differences. In particular, it may be relevant to look at classic effects of perceptual gain control (Hillyard, Vogel, & Luck, 1998) in ASD, because precision is thought to rely on gain control of the output of neural units representing the perceptual prediction errors (Friston, 2009). Foss-Feig, Tadin, Schauder, and Cascio (2013) very recently found that detection of motion direction of a single clearly visible grating can be done based on significantly shorter presentation times in ASD compared to controls. The improvement was not present for low contrast gratings, for which gain control is negligible. Hence, it seems it is caused by reduced contrast saturation of high contrast gratings in ASD, consistent with a deficient gain control.

2.4.2.2 *Local versus global perception*

A common paradigm to study a more locally focused processing style in ASD is to examine their susceptibility to visual illusions.

Overall, these studies yielded mixed to positive effects. While some authors did not find a difference in performance for ASD (C. Brown, Gruber, Boucher, Rippon, & Brock, 2005; Rouse, Donnelly, Hadwin, & Brown, 2004), most others showed a diminished illusion susceptibility in ASD (e. g., Bölte, Holtmann, Poustka, Scheurich, & Schmidt, 2007; Mitchell, Mottron, Soulieres, & Ropar, 2010). This diminished susceptibility has been taken to imply that individuals with ASD are, in general, less influenced by contextual or prior information, remaining closer to the actual sensory input, an idea that is perfectly consistent with HIPPEA. For instance, when Ropar and Mitchell (2002) asked participants to estimate the shape of an illuminated disc presented at a slanted angle in a darkened room, control participants reported a more circular shape (closer to the inferred distal stimulus, discounting the slant), while participants with ASD reported a more elliptic shape (closer to the proximal stimulus, not discounting the slant).

The global-local processing issue is standardly studied using the block design task and the embedded figures task. The first study showed enhanced performance in both of these tasks in individuals with ASD (Shah & Frith, 1993), which was interpreted as evidence for reduced interference by the automatic processing of the global level. Later studies, however, yielded mixed results (e. g., Bölte, Hubl, Dierks, Holtmann, & Poustka, 2008; Ropar & Mitchell, 2001). Collectively, these results point to a difference in degree of efficiency or ease with which the task is performed, rather than a discrete performance difference (Van der Hallen et al., 2014).

Another much researched area of visual abilities in ASD is the perception of motion. A study by Bertone, Mottron, Jelenic, and Faubert (2003) revealed intact first-order (luminance-defined) motion processing but a deficit in second-order (texture-defined) motion processing. Motion coherence studies, in which observers have to track the presence or direction of coherently moving (luminance-defined) dots among differing proportions of randomly moving dots, generally yielded higher motion coherence thresholds in individuals with ASD (e. g., Milne et al., 2002; Pellicano, Gibson, Maybery, Durkin, & Badcock, 2005; Spencer et al., 2000), although there are also exceptions (de Jonge et al., 2007; Del Viva, Igliozzi, Tancredi, & Brizzolara, 2006; Saygin, Cook, & Blakemore, 2010). A recent study may explain this inconsistency (Robertson, Martin, Baker, & Baron-Cohen, 2012) by reporting a deficit in perception of motion coherence at short exposure durations which reduces with increasing exposure durations.

The finding that added noise is especially detrimental for global motion perception in ASD is one that follows directly from HIPPEA.

Distinguishing noise and signal is particularly important in these paradigms. As explained before, people with ASD attribute unduly high value to noise that is unlikely to repeat, in an attempt to properly fit the input. Global motion will more readily 'break' down for them, because they end up with errors that are too important to fit with an abstracted, global pattern. When the noise is absent, as in the plaid motion stimuli in Vandenbroucke, Scholte, van Engeland, Lamme, and Kemner (2008), global motion perception seems to be intact in ASD.

Research with bistable figures suggests that people with ASD can generate and maintain top-down predictions, because when guided to do so, they easily succeed in making the different interpretations of ambiguous figures (Ropar, Mitchell, & Ackroyd, 2003). However, we would advise to use binocular rivalry in future studies (rather than the pen-and-paper type face-vase or duck-rabbit tests used so far) because it has been proposed to be explained by predictive coding (Hohwy, Roepstorff, & Friston, 2008). Indeed input related to the suppressed image in binocular rivalry can be considered prediction error, because it is unexplained by the currently dominant percept. Only two studies have been performed so far with one showing unaltered binocular rivalry in ASD (Said, Egan, Minshew, Behrmann, & Heeger, 2013) and the other finding lower switch rates and more mixed percepts (Robertson, Kravitz, Freyberg, Baron-Cohen, & Baker, 2013). Mixed percepts could be the preferred way to minimize prediction errors in ASD, i.e. less explaining away through higher level constructs and hence staying 'closer to the input'. Note that care should be put into finding the right stimuli for use in autism, since availability of top-down templates evidently also influences rivalry. For example, the first study uses gratings while the second uses familiar objects. The less familiar (or semantically high-level) the better for use in ASD probably, at least when the focus is really on switching dynamics. Future binocular rivalry studies in ASD should specifically look at mixed percepts and fusion, because this the expected result according to predictive coding if precise prediction errors are present for both 'hypotheses' (Hohwy et al., 2008). Another yet to be tested prediction from HIPPEA would be that adding noise (prediction error) to the input has a stronger effect on the breaking of one percept (and possibly inducing a switch) in ASD compared to controls.

2.4.2.3 *Face and speech perception*

Face and speech perception are crucial for smooth and successful social interactions, and therefore prominent targets of ASD research. Deep difficulties here can go a long way in explaining communi-

cation problems so central in ASD. Interestingly, face and speech perception are also prime examples of the hierarchical 'analysis by synthesis' approach inherent to predictive coding. Normally this would provide inferences on high-level semantic sources of incoming sensory information (a generative model) that can cascade into multiple levels of predictions for activity in regions below, suppressing or explaining away new input, as long as it is sufficiently well-predicted. Yet, what is sufficient has to be learned as well (meta-learning), given the requirements of speech or face understanding. If precision of prediction errors is invariably high, individuals with ASD will have more difficulties in 'abstracting away' the short-term, contingent, low-level features of the stimuli. Behaviorally, this is expressed in individuals with ASD as a superior access to the underlying low-level visual or acoustic representations. On the other hand, they cannot fully exploit the higher-level predictions. This disadvantage is particularly felt in naturalistic face or speech-in-noise perception. The problem of which variations to encode in a given situation and which to disregard, comes to the forefront in both speech and faces, which is exactly one of the problems for individuals with ASD, according to HIPPEA. The brain does not only have to pick up and learn small auditory or visual differences, it also has to learn which ones are informative, in the sense of predictive for different kinds of social goals, and which differences to discard.

In speech, invariant phonetic cues are embedded within a variety of acoustic cues (e. g., fundamental frequency, accent, intonation, timbre, etc.) and can only be extracted by integrating and interpolating information, a process which is supported by higher-level linguistic guidance through phonotactic, semantic and syntactic constraints (predictions). During development very young TD children learn to generalize consonants, vowels and words across voices (e. g., of different gender), disregarding irrelevant absolute pitch cues in favor of more complex relative distances. However, in ASD we see increased access to fine-grained acoustic features of complex sounds in ASD (e. g., disembedding tones from musical chords; Heaton, 2003; Mottron, Peretz, & Menard, 2000) and superior perceptual processing of acoustic features of speech (e. g., Heaton, Hudry, Ludlow, & Hill, 2008; Jarvinen-Pasley, Wallace, Ramus, Happe, & Heaton, 2008). Consistent with HIPPEA, it has been suggested that these individuals generate overly specific categories of sounds that impedes learning of higher-level abstract patterns (Crespi, 2013) needed for speech development. Early developmental language delays as well as broader linguistic impairments later in life are indeed prevalent in the individuals who show superior acoustic processing of pitch (Bonnel et al., 2010; Jones et al., 2009).

Additionally, noise with similar characteristics as the signal (speech), substantially hinders performance in ASD (E. G. Smith & Bennetto, 2007), because these 'errors' are not easily ignored.

A similar challenge is posed by faces, characterized by a very high intra-class similarity, with small and rather subtle differences amongst many dimensions distinguishing two human faces from each other. Countless transformations of an individual face amongst several dimensions should be ignored. A face has to be recognized despite variability in, for instance, lighting conditions, face orientation, changeable facial features (e. g., facial hair, spots, wrinkles, freckles), and extra-facial features (e. g., hair style, hats). Due to their meta-learning problems, individuals with ASD may fail to make this distinction between relevant and irrelevant variability, and hence get lost in non-functional characteristics. This may explain their poorer face memory and their face identity recognition problems (Weigelt, Koldewyn, & Kanwisher, 2012).

Paralleling evidence on global-local processing in general, there is no strong evidence for a reduced global or enhanced local face processing style in ASD. For example, no reduced face inversion effect, no attenuated composite face illusion, no diminished part-whole effect, and no decreased susceptibility to the Thatcher illusion was found in ASD (for a review, see Weigelt et al., 2012). More implicit measures, which are less prone to compensatory strategies, do find differences in face processing, contrary to most behavioral studies. For example, children with ASD fail to show the typical longer looking times (van der Geest, Kemner, Camfferman, Verbaten, & van Engeland, 2002) and the typical larger pupil dilation (Falck-Ytter, 2008) for upright than for inverted faces. Moreover, ERP studies demonstrated that the typical differential response to upright versus inverted faces is not present in adults with ASD (McPartland, Dawson, Webb, Panagiotides, & Carver, 2004; Webb et al., 2012). These findings all point to less efficient face processing, because selection and emphasis of predictive cues is missing, throwing face perception back to processes similar to those used for inverted faces. It also fits with HIPPEA that when global face processing deficits are found, they will disappear if participants with ASD are explicitly cued (e. g., "look at the eyes"), as shown by López, Donnelly, Hadwin, and Leekam (2004).

Studies finding disturbed formation of face prototypes in ASD may also confirm our account (Gastgeb, Rump, Best, Minshew, & Strauss, 2009; Gastgeb, Wilkinson, Minshew, & Strauss, 2011). Typically, forming a face prototype requires the use of the central tendency in all encountered exemplar-faces to arrive at an implicit, average representation ignoring the within-category variability (Valen-

tine, 1991). In ASD, however, the emergence of a familiar, broad face prototype will not occur automatically. For categorization to work, new instances have to be recognized as similar to previously experienced examples. The chronically high precision of prediction errors hinders this ability by over-emphasizing the extent to which new input deviates from previous examples or learned templates. Consistently, Plaisted, O'Riordan, and Baron-Cohen (1998b) found that high-functioning adults with ASD learned to discriminate between configurations of colored disks to higher levels of accuracy than controls. However, when tested with slightly different exemplars of the same overall configurations, normal controls showed transfer from the learned exemplars to the novel ones, while individuals with ASD did not. As a result, individuals with ASD may be slower at categorization learning (e. g., Klinger & Dawson, 2001; Soulieres et al., 2011) and they may be less spontaneously extracting a prototype from a series of exemplars (e. g., Gastgeb, Dundas, Minshew, & Strauss, 2012; Vladusich, Olu-Lafe, Kim, Tager-Flusberg, & Grossberg, 2010).

Finally, impaired formation of a familiar, broad face prototype can also be seen in the reduced face adaption aftereffects (Pellicano, Jeffery, Burr, & Rhodes, 2007; Rutherford, Troubridge, & Walsh, 2012, e. g.,). Though these findings may mean that perception is less influenced by prior knowledge (in this case the shifted prototype) (Pellicano & Burr, 2012), we would propose that it is the consequence of an abnormal updating of representations (prototypes). An adapting exemplar may not update the main prototype, because it contains important enough differences for individuals with ASD, to deserve creation of a novel, narrow prototype. Future studies of lower-level feature adaptation, currently lacking in ASD, may be able to resolve this debate.

2.4.2.4 *Mismatch negativity*

While the predictive coding account has originally been conceptualized in the visual domain, a growing number of studies has also investigated predictive coding phenomena in the auditory modality (Arnal & Giraud, 2012; Winkler, Denham, & Nelken, 2009). In this regard, auditory mismatch negativity (MMN) research has been particularly informative. Here, presentation of an unexpected oddball stimulus within a sequence of repeated predictable stimuli, elicits a novelty response in the event-related potential. Originally, the MMN was interpreted as reflecting change detection on the basis of a passive bottom-up process of adaptation to the repeated stimuli (May & Tiitinen, 2010). Recent evidence, however, has shown that the MMN does not reflect release of repetition suppression, but

is the result of a violated prediction rather than a physical stimulus change (e. g., Todorovic, van Ede, Maris, & de Lange, 2011; Wacongne, Changeux, & Dehaene, 2012). A series of studies further made plausible that the perceptual cortex indeed implements a hierarchy of predictions and prediction errors, with repetition suppression attenuating neural responses in a very early time window (40-60 ms), stimulus expectation on the basis of unconscious local predictions attenuating the intermediate stage of processing (100-200 ms, i.e., the typical MMN which originates in sensory areas), and stimulus expectations on the basis of more global, integrative and conscious predictions modulating the later P3b novelty response (300-600 ms, originating from a broader frontoparietal predictive network) (Todorovic & de Lange, 2012; Wacongne et al., 2011). Regarding the MMN, a number of studies observed larger amplitudes and/or earlier latencies to infrequent pitch changes in tones and vowels in ASD relative to TD controls, thus suggestive of hypersensitivity and superior recognition of pitch change (e. g., Ferri et al., 2003; Gomot, Giard, Adrien, Barthelemy, & Bruneau, 2002; Lepisto et al., 2005 but see Dunn, Gomes, & Gravel, 2008). Interestingly, Gomot et al. (2011) showed that these electrophysiological abnormalities were significantly more pronounced in children who displayed greater difficulties in tolerating change. The MMN response to infrequent phonemic changes in vowels or consonants, however, is typically smaller and/or delayed in ASD, thus suggestive of impaired recognition of the more global phonetic characteristics of speech (e. g., Kujala, Lepisto, Nieminen-von Wendt, Naatanen, & Naatanen, 2005; Lepisto et al., 2006). Finally, the later P3b component, presumably characterizing more global and integrative violations of expectations, exhibits smaller amplitudes in ASD relative to controls (e. g., G. Dawson, Finley, Phillips, Galpert, & Lewy, 1988; Kemner, Verbaten, Cuperus, Camfferman, & van Engeland, 1995). Comparing neurophysiological findings pertaining to MMN versus P3b processing suggests that the brains of individuals with ASD are tuned to register low-level local changes in transition probabilities (cf. enhanced and earlier MMN sensory responses towards simple stimuli), but have difficulty picking up changes in the broader fronto-parietal predictive system which is tuned towards more global, higher-level patterns. This is at least compatible with the view that increased low-level precision hinders the formation of appropriate predictions higher up.

2.4.3 *Savant skills*

'Autistic savants' are individuals with ASD with co-occurring excellence in an isolated skill, i.e., an 'island of genius' which contrasts with the individual's general lower-than-average abilities. Savantism has been identified in a wide range of neurological and neurodevelopmental disorders, but is most frequently reported in ASD. Savant skills are estimated to be present in one out of 10 autistic individuals, with males outnumbering females (approx. 6:1) (Howlin, Goode, Hutton, & Rutter, 2009; Treffert, 2009). Savant skills usually fall within one of five general categories, i.e, musical abilities, calendar calculating, mathematics, art and mechanical or spatial skills (Treffert, 2009). Although the savant skill of an individual may evolve over the years, the skill should not fade or disappear over time, but remain a peak in performance.

Several scholars attempted to explain the mechanism behind the savant skills. Plaisted (2001) suggested a reduced ability to process similarity at the perceptual and attentional level which results in a reduced tendency to generalize information. Baron-Cohen (2006) postulated an increased drive to construct or analyze, which he referred to as 'hyper-systemizing'. The alleged adaptive function of the systemizing mechanism is to serve as a law-detector and a change-predicting mechanism. He argues that people with ASD prefer either no change, or systems which change in highly lawful or predictable ways (i.e, systems with rule-bound change, such as mathematics, physics, objects that spin or recur, music, machines, collections), and why they become disabled or 'change-resistant' when faced with systems characterized by 'complex' change (such as social interaction). Mottron et al. (2006) and Mottron et al. (2013) emphasized the putative role of enhanced feed-forward low-level perception and suggested that individuals with ASD have a developmental predisposition to 'veridical mapping' of data and information. Although these accounts provide insight into the origin of such a skill, HIPPEA makes more specific claims about the underlying mechanisms.

Our predictive coding approach explains why similarity is not processed in the same way in ASD, consistent with Plaisted (2001). It also elucidates why complex change is challenging (Baron-Cohen, 2006): this is where meta-learning should lead to distinguishing mere noise from actual environmental changes. Finally, the veridical mapping can also be seen as a consequence of the constant drive to reduce even irrelevant prediction errors (Mottron et al., 2013). While predictions shaped by noise and irrelevant details will often result in impaired or slow processing, doing this for a specific, lim-

ited topic of interest can be quite possible and, above all, rewarding. Developing such a skill becomes extrinsically motivating (e. g., getting praise and attention) but also intrinsically, as making successful predictions in this particular domain will result in feelings of reward, and the notion that the generally unpredictable world is more controllable. For example, phone numbers have an exact but arbitrary mapping (Mottron et al., 2013). All known examples of savant skills, for instance, 3D drawings or musical play from memory, combine two factors: an exquisite discriminative sensory ability and an exceptional (rote) memory capacity (A. L. Hill, 1978; Treffert, 2009). The first is a general feature of ASD, we would argue, originating from high precision low-level prediction errors. A lack of abstraction is actually an advantage here. Clearly, this discriminative ability can only fully be put to use in the case of high memory capacity. This may be the feature that is specific to savants, but even then resource constraints may seriously limit the savant domain.

2.5 SENSORIMOTOR ABILITIES AND A SENSE OF SELF

Within the predictive coding theory, actions also entail predictions, namely of their proprioceptive and exteroceptive consequences. According to M. J. Edwards, Adams, Brown, Pareés, and Friston (2012, p. 3498), movement is defined by "what we want to see (or feel), rather than what we want to do". In this view, actions can be regarded as being aimed at fulfilling predictions (reducing prediction errors) of perceptual input. Several ASD symptoms can be readily interpreted from this perspective. Given that actions generate prediction errors, those actions that reduce these prediction errors to extreme minima should be preferred. Accordingly, some of the most characterizing symptoms in ASD are the stereotypical, repetitive (predictable) behavior patterns (Turner, 1999). These patterns establish controllable and thus very predictable proprioceptive (kinesthetic) feedback, that helps individuals with ASD to better cope with their environment (Ornitz, 1974). In a similar vein, the repetitive handling of lighting and spinning objects, and the repetitive tactile self-stimulation can be regarded as manners of creating a predictive environment to reduce and cope with prediction error. Especially unpredictable surroundings may be expected to elicit this kind of behavior, with the aim of reestablishing predictability and reducing stress (see Section 2.6). Ornitz (1974, p. 204) observes that "In their spontaneous activity autistic children are continually spirting, twirling, flicking, tapping, or rubbing objects. Furthermore, they repetitively flap, writhe, wiggle, or oscillate their extremities while regarding them intently." This latter part is significant because

it indicates that while TD children might progress to more complex kinds of 'play' (learning), children with ASD continue to be engaged in and learn from these simpler patterns.

According to HIPPEA, atypical behavior has the aim of regulating excessive amounts of prediction errors. At first sight, this seems very similar to the explanation invoked by the EPF theory, namely reducing excessive perceptual input (Mottron et al., 2006). However, in our view, individuals with ASD only aim to reduce that part of the perceptual input that cannot be predicted, and moreover actively attempt to create predictability to compensate. Interestingly, reports of autistic children screaming all day, despite being hypersensitive to noise themselves, might be understood as a way of dealing with prediction errors by making the sensory environment more predictable. The active desire for predictable sensory experience is brought even more clearly into light by Temple Grandin, an autistic woman who built a mechanic body squeeze machine, because she liked the feeling of being touched and hugged, but wanted it to be perfectly controlled (i.e., predictable) instead of the unpredictable overstimulating human touch (Edelson, Edelson, Kerr, & Grandin, 1999; Grandin, 1992). In a similar vein, the 'high systemizing' concept used by Baron-Cohen, Ashwin, Ashwin, Tavassoli, and Chakrabarti (2009) to characterize the cognitive style of individuals with ASD can underscore that predictable patterns are formed and are important in their minds. The obsession with regularity can be seen as borne of an overweighing of deviations.

The sense of self and of agency has also been related to (interoceptive) predictive coding (Apps & Tsakiris, 2014; Seth, Suzuki, & Critchley, 2012). It is through the tightly cross-modally correlated proprioceptive, tactile and visual input of self-induced movements that we construct the sense of a self that acts in the world. The high-level concept of the self is the most plausible prediction explaining low-level regularities in cross-modal input. This view of the emergence of the self via the observed correlations between proprioceptive, tactile and visual modalities can also explain why artificially created correlations can create the illusion that extra-corporeal objects are part of our own body (e. g., rubber hand illusion; Apps & Tsakiris, 2014; Botvinick & Cohen, 1998). Awareness of self and body as distinct from the world is thus dependent upon a certain degree of tolerance derived from the active, successful suppression of interoceptive prediction errors (Seth et al., 2012). The presence of repetitive, stereotyped movements in ASD during early development suggests that an abnormally large amount of correlated input is needed to establish a sense of self as separated from the surroundings (see also Brincker & Torres, 2013).

Two recent studies using the rubber hand illusion, an illusion of perceived arm position induced by correlated (synchronized) stimulation (Palmer et al., 2013; Paton, Hohwy, & Enticott, 2012), support this view. Both in individuals with ASD and in those with high but nonclinical ASD traits, the consequences of experiencing the illusion (on drift and movement) were reduced. A higher estimated precision of prediction errors may indeed lead to a reduced illusory percept, requiring more tightly correlated input (than is usually provided in this rubber hand procedures) for the illusory percept to fully establish itself. More generally, motor coordination problems, often noted in ASD (Fournier, Hass, Naik, Lodha, & Cauraugh, 2010; M. L. Matson, Matson, & Beighley, 2011), may be another consequence of over-precision of movement prediction errors in contexts that actually have a considerable amount of uncertainty (Palmer et al., 2013).

The finding that the repetitive, self-focused behaviors often decrease during development (Richler, Huerta, Bishop, & Lord, 2010) suggests that extensive exposure may eventually lead to a more stable sense of self. However, the typical insistence on sameness (Kanner, 1943) remains or increases with age, indicating that exteroceptive prediction errors generally remain precise. This insistence on routine or rituals, and resistance to trivial changes in the surroundings, again demonstrate that children with ASD do develop clear predictions on what should happen next in the current situation, in contrast to theories positing a uniformly weaker application of predictions in ASD (Pellicano & Burr, 2012). Therefore, insistence on sameness may be considered a hallmark of HIPPEA: It signals a clear grasp (prediction) on how the world should behave, while assigning too much importance to incidental changes.

2.6 CHRONIC UNPREDICTABILITY AND ITS AFFECTIVE CONSEQUENCES

One of the most prominent clinical observations in individuals with ASD is their unusual reactivity to sensory stimuli. Numerous clinical and personal reports describe the presence of both hyper- and hyposensitivity to sensory stimulation. Hypersensitivity has been described in various modalities (Blakemore et al., 2006; Kern et al., 2006; Khalfa et al., 2004). Enhanced sensitivity to loud and unexpected sounds is particularly evident in children with ASD (e. g., Grandin, 1995; Tomchek & Dunn, 2007), and appears to decrease with age, with adults with ASD becoming more similar to TD adults (Kern et al., 2006). Yet, feelings of stimulus overload and hypersensitivity to noise are also common in adults on the autistic spectrum

(in particular in social situations, like receptions or parties), and can cause great distress and anxiety. Enhanced sensitivity to visual stimuli is less common in ASD, but does occur, e. g., under the form of enhanced discomfort to bright light (Kern et al., 2001). When the gain of the neural units representing the prediction errors is fixed at a high level, it is easy to see that hyper-sensitivity becomes very likely, especially for unexpected input, as is the case in ASD. Overweighting of irrelevant prediction errors causes sensory overload.

Seeing that unpredictability is at the core of the sensory overload, we can also attempt to explain its negative affective impact. Uncertainty has long been identified as a factor that intensifies stress and anxiety (Herry et al., 2007; Miller, 1981). In addition to leading to increased stress and anxiety, persistent significant prediction errors may actually by themselves generate negative affect (Huron, 2006; Van de Cruys & Wagemans, 2011b). When predictions are invoked, there is actually something at stake, namely the success of current internal models of the environment. When prediction errors signal the need for extra resources, aimed at updating the internal model, they may have negative affective value. For example, supposedly neutral perceptual prediction errors activate the habenula, a region known to code prediction errors of negative valence (Schiffer, Ahlheim, Wurm, & Schubotz, 2012; Schiffer & Schubotz, 2011). Originating from the cognitive dissonance tradition, recent frameworks in social psychology center precisely on the link between expectation violation (or uncertainty) and anxiety, with much of human cognition and behavior interpreted as efforts to reestablish a coherent, predictable world model (Hirsh, Mar, & Peterson, 2012; Proulx, Inzlicht, & Harmon-Jones, 2012).

The taxing, negative experience described in ASD as sensory overload or oversensitivity is, according to HIPPEA, a logical consequence of a brain continuously signaling that prediction errors merit the recruitment of more resources for learning. The proactive (predictive) investment of the system makes this a particularly aversive experience. Conversely, making progress in predicting the world (reducing prediction errors) may genuinely feel rewarding. Note that not the static state of low prediction error but rather the transition (change) from a state of high prediction errors to a state of low errors may induce positive affect (Joffily & Coricelli, 2013; Oudeyer et al., 2010; Van de Cruys & Wagemans, 2011b). This kind of reward arguably is the driving force for further exploration and learning (cf. Section 2.3). However, difficulties in estimating where predictive progress can be made could largely rob a person from experiencing this type of reward, with detrimental implications for intrinsic motivation. Indeed, problems in general motivation and exploration are

reported in ASD (Koegel & Mentis, 1985; Ozonoff et al., 2008), from very early on in development (Zwaigenbaum et al., 2005).

The combination of increased uncertainty-related anxiety and decreased reward of exploration may have particularly incapacitating and far-reaching effects in the longer term. We already referred to learned helplessness to indicate the anxious avoidance and lack of motivation caused by repeated frustration in predicting one's surroundings. By caregivers this may be interpreted as hypo-reactivity (Ben-Sasson et al., 2009; Tomchek & Dunn, 2007). Social interactions arguably suffer most from this lack of motivation (Chevallier, Kohls, Troiani, Brodkin, & Schultz, 2012), with obvious consequences with regard to the willingness to engage in social relations. We do not consider social motivation problems to be the origin of ASD, but our account agrees with social motivation theories (Chevallier et al., 2012) that this is an important aggravating factor in the syndrome. Indeed, social interactions are not perceived that enjoyable or rewarding in individuals with ASD (Chevallier et al., 2012). Unsurprisingly, a lot of interventions focus on increasing the reward of social interactions. If social situations are avoided from early on in life, the number of social learning experiences decreases, and so, in a vicious circle, even more social impairments ensue.

Taken together, these factors arguably make individuals with ASD more vulnerable to mood and anxiety problems, which are indeed overrepresented in ASD (Kim, Szatmari, Bryson, Streiner, & Wilson, 2000). Hence, mood problems, anxiety and anxious avoidance should in our view be considered as secondary symptoms, originating from accumulated experience with (irreducible) prediction errors, and from repeated frustration in learning. Consistent with this, anxiety and mood problems seem to increase during childhood in ASD (Kim et al., 2000).

2.7 SOCIAL FUNCTIONING

Social interaction problems are amongst the first described symptoms of ASD (Asperger, 1944; Kanner, 1943) and are crucial pieces in the DSM-classification (American Psychiatric Association, 2013). Social impairments stand out strongly in the clinical phenotype (demonstrated by the existence of ASD questionnaires focusing only on the social symptoms; e. g., Constantino, 2002), and retrospective studies often report early signals in the social domain (Volkmar, Chawarska, & Klin, 2005). The phenomenal and clinical prominence of social deficits spurred a wealth of evidence on social impairments. Therefore, a central challenge for core information processing dys-

function theories of ASD is to explain why abnormalities manifest themselves most clearly in the social domain.

What sets social situations apart from non-social situations? Or better: what distinguishes social tasks in the lab from the tasks used for other (lower-level) domains? Like Simmons et al. (2009), we wonder whether *social* may just be a synonym of *complex* here. However, our approach allows us to pinpoint exactly what this complexity may entail with regard to the difficulties in ASD. Most ingredients have been provided in the previous sections, but in the social domain they come together and are expressed to the fullest.

2.7.1 *Social complexity*

Our brief overview of face and speech processing impairments in ASD did not strongly speak for a special status of faces or speech as such. Here too, we do not want to treat social judgements differently from other processing. It is more fitting, we argue, to view them as just another kind of inference, in this case inference about other people's emotions or intentions from their facial expressions, gaze, bodily postures, etc. (Hohwy & Palmer, 2014; Zaki, 2013). Therefore, the same mix of accidental uncertainty and informative changes determines the social problems in ASD. No two social scenarios are identical. Numerous accidental properties in the rich social environment are mostly uninformative and should be ignored. This is ideally what tuning down precision should accomplish. Individuals should (meta-)learn which aspects are informative and which are irrelevant to the social rules governing the current situation. This is particularly difficult when these noisy social contingencies are changing and context-bound, which they mostly are (Barrett, Mesquita, & Gendron, 2011). There is rarely a one-to-one mapping between social signals and their meaning. For example, happiness can be expressed with an obvious loud laughter, but an enigmatic Mona Lisa smile is possible too. A similar laugh can signify consent (humor) or rejection (irony). Subcultures (e. g., youth culture) invent new meanings for old signals (e. g., words) or new signals for old meanings. In addition, low-level input can be dramatically different while the same social rules apply. Instead of flexibly adjusting the precision of prediction errors based on previous and current experiences, individuals with ASD will get flooded by the wealth of available information in a social situation.

Generalizing what we said about face perception, people with ASD fail to discriminate between informative and irrelevant properties when making social judgments (cf. the lack of autonomous selection in attention). The result is that social information does not

seem to be particularly salient for them, or at least not more so than non-social stimuli. This deficit is most clearly illustrated by eye-movement studies. Individuals with ASD show a reduced attention to faces, but more attention towards bodies and objects in the background of a social scene (e. g., Klin, Jones, Schultz, Volkmar, & Cohen, 2002; Rice, Moriuchi, Jones, & Klin, 2012). Within faces too, differences in information selection are noticeable. They also do not seem to have learned the typical informativeness of the eyes region, crucial for face and emotion recognition. Instead, studies reveal a bias for the mouth region and scanning patterns towards the outer face characteristics (such as hair; Harms, Martin, & Wallace, 2010). From early childhood on, children with ASD do not show the usual preference for social stimuli (Klin, 1991, 1992). Two year old children with ASD rather attend to non-social physical contingencies instead of socially relevant biological motion (Klin, Lin, Gorrindo, Ramsay, & Jones, 2009). We think this should be explained by the steadier, lower-level predictability of the former.

The fact that the atypical viewing patterns and the emotion recognition deficits are most apparent when using complex stimulus material (Chevallier et al., 2012; Harms et al., 2010) also speaks for our hypothesis. While the distinction between relevant and irrelevant information may be rather clear-cut in simple social stimuli (e. g., isolated, well-controlled and re-used faces), using ecologically valid stimuli (e. g., noisy, dynamic social scenes) implies more competition from distracting (irrelevant) information.

Gradually, TD children form 'social scripts': abstracted and broadly applicable knowledge structures, representing an ordered sequence of actions, causes and consequences within a certain social context (e. g., making friends). In children with ASD this capacity to generate adequate social scripts is found to be impaired (Loth, Happe, & Gomez, 2010). It is easy to see that indiscriminate precision of social and non-social cues results in narrow and specific social scripts (e. g., making friends *when* I'm playing soccer), wrought with spurious, concrete features. Interventions that try to remedy social script deficits, select and describe the relevant cues for a given script, linking it with possible appropriate responses (for a meta-analysis, see Reynhout & Carter, 2011).

2.7.2 *Multisensory integration*

Adequate social understanding heavily relies on integration of multiple sources of information, both within modality and across modalities. The same facial expressions can receive completely opposite meanings depending on the bodily context in which they ap-

pear (Aviezer, Trope, & Todorov, 2012). In other situations different modalities provide complementary information, to be used to figure out emotions and intentions from face-to-face-communication. In such cases, additional information of another modality helps the interpretation. For instance, visual articulatory information aids speech perception, especially under noisy circumstances. Again we note that uncertainty of the different sources has to be taken into account in order to determine which information should have more saying in the eventual social judgment. Indeed, this can be formalized using Bayes' theorem (Zaki, 2013), which is already widely used in (non-social) perceptual cue integration studies. For optimal inference, the expected uncertainty (precision) of the different sensory sources should determine differential reliance (weight) on those sources.

Individuals with ASD are known to have difficulties with such multisensory integration (Iarocci, Burack, Shore, Mottron, & Enns, 2006), for instance with the detection of inter-modal correspondence of facial and vocal affect (e. g., Loveland et al., 1995). If precision is fixed at a similarly high level for all sources, as HIPPEA maintains, optimal integration will not take place, because all cues, even redundant or very uncertain ones, will be weighed equally. Moreover, the spatiotemporal contiguity of two inputs required to perceive them as belonging to the same distal cause would be more strictly defined for people with high precision. Any minor spatiotemporal mismatch between two cues (e. g.visual-auditory in the ventriloquist effect or visual-haptic in the rubber hand illusion) will render it more likely that these will be experienced as distinct unimodal events rather than an integrated, multimodal event (Palmer et al., 2013). The attenuated McGurk effect found in ASD could similarly be explained (Mongillo et al., 2008; Taylor, Isaac, & Milne, 2010).

2.7.3 *Mentalizing*

'Theory of Mind' or 'mentalizing' refers to the ability to read the (facial) expressions of other people, to understand their feelings, intentions, wishes and thoughts, and to use this —mostly implicit— knowledge to understand another individual's actions and guide one's own actions (Premack & Woodruff, 1978). A vast amount of research in ASD has focused on the Theory-of-Mind problems in individuals with ASD, arguing that individuals with ASD have difficulties in placing themselves into the mental world of others and themselves, and sometimes described as 'mindblindness' (e. g., Baron-Cohen, 2001; Frith, 2003). The discovery of mirror neurons (in monkeys) that are active both during the action observation or imag-

ination (offline processing), but also during the online execution of an action (e. g., Kohler et al., 2002), led to the conjecture that action-understanding and even mentalizing crucially rely on this class of neurons. All discussions on the precise role and distribution of mirror neurons in the brain aside, this finding conclusively showed that action execution and action perception are closely intertwined. Predictive coding offers a new perspective on the implementation of goal and intention inference in the mirror system (e. g., Friston, Lawson, & Frith, 2013; Kilner, Friston, & Frith, 2007; Koster-Hale & Saxe, 2013). As mentioned before, actions could be conceived of as a series of hierarchical predictions (Hamilton, Brindley, & Frith, 2007), going from longer-term intentions and goals (e. g., to splash water in your friends' face) over short-term goals (e. g., to grasp a glass of water) and motor plans (movement sequences), down to the muscle commands and kinesthetics. At all levels, predictions will be matched with input, resulting in prediction errors that drive and guide proper action execution. Importantly, in predictive coding this same hierarchical model that is used for forward action generation also serves inverse inference: figuring out goals from observed actions (Kilner et al., 2007). Observed actions will both automatically generate expectations on the kinematics and muscle activation linked to it, and create discrepancies that can only be explained away by inferring an appropriate intention on the highest levels. How can this system distinguish then between own actions and another's actions? Put differently, observed action creates prediction errors because motor plans and goals are generated, while muscle and kinesthetics are inactive. How does the brain avoid automatically executing (mimicking) observed actions to reduce those low-level prediction errors?

The assumed mechanism is, again, precision (Clark, 2013a). The prediction errors have a high expected precision, which makes sure actions you initiate yourself are properly executed. These prediction errors will be suppressed by your own accurate predictions (goals), often inciting a sense of confidence or agency (see Section 2.5). For action observation however, estimated precision of motor prediction errors should be tuned down, such that they receive low weight and the thrust of processing moves to higher-level inference of goals and intentions. In this way, precision becomes the mechanism that allows organisms to exploit the learned hierarchical models for action execution, also for mentalizing and offline planning (Clark, 2013a).

Following this reasoning, a deficit in the flexible tuning of precision of prediction errors, resulting in an overly high estimation of precision, as HIPPEA assumes to be the case in ASD, may give rise to a couple of related problems. First, it may contribute to offline (mo-

tor) planning problems (Booth, Charlton, Hughes, & Happe, 2003; Hughes, 1996), with high precision preventing individuals to transcend the immediate input, as noted earlier. Second, failure to lower the precision of low-level prediction errors during action observation, may automatically lead to precise proprioceptive prediction errors, because the action is not executed. A possible strategy to reduce these errors is the mimicking of (formal aspects of) others' behavior. Indeed, hyper-imitation of formal aspects of behavior (Bird, Leighton, Press, & Heyes, 2007; Spengler, Bird, & Brass, 2010), and echolalia and echopraxia, the automatic copying of others' speech or behavior occur more frequently in the ASD population. We are cautious in pointing to this possibility, because precision of motor errors may be determined by a different neurotransmitter (dopamine) than perceptual errors (see Section 2.8) and not every child with ASD shows this automatic mimicking.

A third possible problem of inflexible tuning of precision, links back to our discussion on visual and auditory perception. We noted there that top-levels of hierarchical models may not get properly build (learned), because processing is stuck in low-level matching due to the high precision of low-level prediction errors. If for motor execution and planning too, individuals with ASD end up with incomplete hierarchical models, they may be unable to reach the higher levels of conceptual inferences of goal and intention. Consequentially, these individuals will experience difficulties in inferring emotions from their *own* bodily states and expressions (Seth et al., 2012). Indeed, alexithymia is often found in ASD, and has recently been shown to better predict poor recognition of emotional expressions than ASD as such (Cook, Brewer, Shah, & Bird, 2013). From the predictive coding standpoint, where one model is used for both emotion recognition and inferring own emotion, this makes a lot of sense. Brain responses related to empathy are also modulated by alexithymia rather than ASD (Bird et al., 2010). If these findings are corroborated, it may turn out the empathy and emotion recognition problems in ASD (see Harms et al., 2010; Uljarevic & Hamilton, 2013, for a meta-analysis and literature review, resp.) are not primary symptoms, but are inherently linked to alexithymia. The processing profile of ASD as we sketched may predispose patients to alexithymia, because high precision interoceptive prediction errors prevents adequate emotional inferences (Seth et al., 2012).

2.8 NEUROBIOLOGICAL UNDERPINNINGS

In the current paper we primarily wanted to articulate the cognitive, computational foundation of our account and its behavioral

consequences. We do want to briefly survey plausible neurobiological underpinnings of the proposed mechanism as well, without giving an exhaustive review of the neurobiology of ASD (e. g., see D. G. Amaral, Schumann, & Nordahl, 2008; Bauman & Kemper, 2005). Clearly, a more systematic, extensive discussion of ASD neurobiology in light of HIPPEA will be needed in the future.

Using HIPPEA, we can tentatively divide neurobiological findings in three parts: first, studies directly targeting the neural regulation of precision, second, studies on the neural basis of models of uncertainty and meta-learning that feed into regulation of precision, and third, downstream consequences of high precision for neural plasticity and connectivity. We will only consider the first two here and leave the last part for later work.

2.8.0.1 *Precision regulation*

In Friston's predictive coding model, precision is regulated by neuromodulators that control the gain of the units representing prediction errors (Friston, Daunizeau, Kilner, & Kiebel, 2010). This gain determines the impact of prediction errors on units that encode the predictions. Neuromodulators such as acetylcholine (ACh) and norepinephrine (NE) are long known to influence attention and learning, so they are likely candidates for this role. In particular, the neuromodulator ACh is assumed to enhance precision of perceptual prediction errors (Friston et al., 2010). Indeed, a pharmacological agent that increases ACh availability in cholinergic synapses increases the event-related response to deviations of predictions (Moran et al., 2013) and attenuates the decrease in activity with repeated stimulation (repetition suppression). However, Yu and Dayan (2005) proposed a different, complementary role of ACh and NE, in which only expected uncertainty, linked to the known stochasticity (lack of reliability) of a predictive relationship, is coded by ACh. NE, on the other hand, tracks unexpected uncertainty, that is, the actual, important changes in the regularities governing the relationships in the world (Duzel & Guitart-Masip, 2013; Payzan-LeNestour, Dunne, Bossaerts, & O'Doherty, 2013). A context-dependent modulation of the balance between these two must ensure that learning is enabled when learning is due (for actual changes).

The findings on (nicotinic) cholinergic signaling in ASD are very inconclusive at this stage, but a few studies report abnormalities (Lam, Aman, & Arnold, 2006), including in the main source of ACh, the basal forebrain (Bauman & Kemper, 1994; Perry et al., 2001). Raised NE signalling in ASD is suggested by elevated blood plasma levels (Lam et al., 2006) and by a tonically high arousal system as shown by a tonically elevated heart rate in autistic children,

with reduced phasic response (Kootz & Cohen, 1981). For pupil size the same pattern has been reported: increased tonic pupil size, and increased latency, smaller constriction amplitude and lower constriction velocity for the pupillary light reflex, compared to TD children (Anderson & Colombo, 2009; Fan, Miles, Takahashi, & Yao, 2009). This is noteworthy, because of the known coupling of pupil size with the NE system, more specifically with activity in the principal source of NE projections, the locus coeruleus (Rajkowski, Kubiak, & Aston-Jones, 1993). Finally, prenatal overstimulation of the β_2-adrenergic receptor by an agonist is associated with increased risk of ASD (Connors et al., 2005).

Hence, available evidence already seems to point to some loss in the dynamic range of ACh and NE neuromodulation, but direct tests await. Pharmacological studies applying an agent that increases central cholinergic signaling should verify whether the ERP or behavioral response to expectation violation is modulated similarly in individuals with and without ASD (cf. Moran et al., 2013). If cholinergic signaling is already at ceiling in ASD, an additional boost of this system may not make a difference. Alternatively, a cholinergic antagonist may, in ASD, lead to 'normal' performance on tasks that benefit from disregarding smaller differences (on which ASD subjects are usually worse). With regard to NE, there may be considerable potential in measuring pupil dynamics in ASD. Nassar et al. (2012) demonstrated that learning dynamics can be tracked by pupil size measurements, suggesting that NE arousal systems indeed can regulate learning. Their predictive inference task required adjusting of precision (learning rate), because predictive relationships changed at certain points ('change points') in the course of the task, as explained before (see Section 2.4.1). Apparently, pupil diameter change is monotonically related to change point probability, where prediction errors should indeed receive high weight. Additionally, average pupil size reflects "uncertainty that arises after change points and signals the need for rapid learning" (Nassar et al., 2012, p. 1043). Recall that this uncertainty was called reducible. If ASD is linked to increased precision of prediction errors across the board, as HIPPEA maintains, this should be apparent both in average learning rate and pupil metrics in this sort of task.

Finally, there is evidence that these neuromodulators can act as metaplastic signals regulating the potential of synapses to undergo activity-dependent long-term potentiation (e. g., Inoue et al., 2013). This provides another link with precision as a meta-learning signal, that needs to be explored more. Indeed, several of the genetic mutations linked to ASD have an important role in the regulation of plasticity (e. g., Delorme et al., 2013; Ebert & Greenberg, 2013;

Hutsler & Zhang, 2010). Relatedly, the valproic acid rat model of ASD shows twice the amount of long term potentiation of controls (Markram & Markram, 2010).

2.8.0.2 *Models of uncertainty*

We emphasized before that precision of prediction errors does not appear out of the blue. The brain builds meta-models, predictions of prediction errors, to estimate precision. These meta-models are formally not that different from regular predictive models assumed to take place across the perceptual hierarchy. Arguably then, these meta-models may be represented also in a distributed manner across the cortex. However, there is evidence that some regions are more involved than others in the processing of uncertainty.

Two regions that are good candidates for this *and* that have recently attracted researchers' interest in ASD are the insula and the anterior cingulate cortex (ACC). Both are thought to be central parts of the so-called salience network, the circuit involved in responding to behaviorally important stimuli and in cognitive control. Indeed, we could replace the somewhat vague term 'salience' with 'precision', because *in se* they have similar intent, namely determining value or relevance of input for behavior and learning. The salience network is closely connected to the motor system, suggesting a role in generating exploratory actions (Rushworth, Behrens, Rudebeck, & Walton, 2007), as we discussed in Section 2.3 on exploration in ASD. Also, it is deemed to be crucial in judging whether to persist in or switch the current attentional set (Dosenbach et al., 2006). Evidently, models of uncertainty in input are vital in such decisions. Finally, the ACC innervates the locus coeruleus-NE system (Aston-Jones & Cohen, 2005), perhaps allowing it to modulate gain (precision) of prediction errors in the (sensory) cortex.

A recent study found hyperactivation in dorsal ACC in response to visual oddbal stimuli in ASD (i.e., infrequently presented, deviant stimulus) (Dichter, Felder, & Bodfish, 2009), consistent with the idea that expectation violations are more salient. Healthy subjects show ACC activity for behaviorally relevant prediction errors (Ide, Shenoy, Yu, & Li, 2013; Metereau & Dreher, 2013). Others have found evidence that the cingulate cortex not only represents the prediction errors, but also performs the computations underlying the adaptive regulation of precision (D. R. Bach, Hulme, Penny, & Dolan, 2011; Behrens et al., 2007).

The insula too is known to be involved in prediction under uncertainty. Activity in ACC and insula is strongly coupled, and critically, this coupling is modulated by prediction errors (Limongi, Sutherland, Zhu, Young, & Habib, 2013). Using a gambling game,

Preuschoff, Quartz, and Bossaerts (2008) showed that activity in the anterior insula can code that part of uncertainty that cannot be reduced, due to the stochasticity of the associations at hand, also called known or expected uncertainty. There is evidence from dynamic causal modeling analyses that anterior insula is the entry point of the salience network and drives ACC activity (Ham, Leff, Boissezon, Joffe, & Sharp, 2013; Limongi et al., 2013). If true, a possible hypothesis is that the insula corrects incoming prediction errors for known stochasticity and thereby helps ACC and further regions to properly attribute salience (precision) of the prediction errors. In any case, insula, ACC and possibly neighboring frontal regions may cooperate to dissect uncertainty, with the aim of estimating where predictive progress can be made and setting precision accordingly (attending to the right aspects of input) (see also Karlsson, Tervo, & Karpova, 2012). When, as is the case for ASD, there is abnormal connectivity and activity of ACC and insula (Di Martino et al., 2009; Uddin & Menon, 2009), this estimation process may go awry, leading to unadaptive and possibly chronically high precision. Much more work is needed because existing neuroimaging work in ASD mostly uses task contrasts (often using faces or other complex stimuli) that are hardly informative in relation to our proposal. Fortunately, the needed paradigms have already been applied in nonclinical participants.

2.9 RELATED APPROACHES

Several important theoretical frameworks of ASD can be usefully compared to ours. Some have been emphasized in previous sections, but here we more closely look at those that were not discussed before and that are most akin to our theorizing, in postulating a broader information-processing account. In the second part of this section, we will address the question whether a unified account is possible at all, in view of the heterogeneous nature of ASD. We will close this section with a discussion of a recent theory of schizophrenia, which is closely related to our theory of ASD.

2.9.1 *Other information processing accounts of ASD*

A straightforward, Bayesian way to conceptualize problems in ASD could be to assume broader (high uncertainty) priors or predictions that therefore have a weaker influence on the outcome of perceptual inference. Indeed, this road has recently been taken by Pellicano and Burr (2012) in a thought-provoking article (for a related approach, see Gomot & Wicker, 2012). These authors argued that this may

cause perceptual outcomes to remain closer to the perceptual input, minimally biased by top-down, prior knowledge, an idea that is consistent with the WCC theory. Hence, this account explains why individuals with ASD may be less susceptible to visual illusions that are caused by prior knowledge or contextual interactions (see Section 2.4.2 on perception). In other words it would, according to the authors, result in a more accurate or 'real' perception.

In addition to spurring an interesting discussion (Brock, 2012; Friston, Lawson, & Frith, 2013; Teufel, Subramaniam, & Fletcher, 2013; Van de Cruys et al., 2013; van Boxtel & Lu, 2013), this stance has been criticized on theoretical and empirical grounds. Teufel et al. (2013) remind us that "a perceptual system that refines sensory information by prior knowledge provides a better estimate of real but hidden causes than perception that is based on the ambiguous sensory information on its own, because the former system exploits all the relevant information available." In this regard, broader priors would lead to less accurate perception because the actual input is always noisy and ambiguous. Even in the case of visual illusions, it is not "priors per se [that] render perception less accurate; rather, it is the application of the wrong prior that leads to the illusory percept" (Teufel et al., 2013). Furthermore, Brock (2012) notes that perception (the posterior) can move closer to perceptual input (likelihood) for two different reasons: either, as Pellicano and Burr (2012) argue, the prior is broader (higher uncertainty, lower precision) *or* the likelihood is sharper (lower uncertainty, higher precision). It should be clear that the proposal in HIPPEA is more akin the second option. Finally, there is evidence that individuals with ASD are very well capable of building precise expectations from experience (see Section 2.4.1). Indeed that may be the reason why they are so perturbed by information that deviates from this information. The problems, we argue, arise because these deviations receive too much salience. Instead of a lack of precision in predictions, there may be a heightened precision of prediction *errors* in ASD.

It is also interesting to distinguish the our view from approaches locating the core problem in ASD in a reduced signal to noise ratio in neural processing (Belmonte et al., 2004; Simmons et al., 2009). While increasing noise usually impairs psychophysical performance, it can improve detection under restricted conditions, a phenomenon called 'stochastic resonance' (Goris, Wagemans, & Wichmann, 2008). Though speculative at this stage, increased internal noise in neural communication may in this way be able to explain both improved performance in a limited number of tasks and impaired performance on more complex, high-level tasks (Simmons et al., 2009). HIPPEA in contrast, does not necessarily assume increased

internal noise in neural signaling, but rather a higher weighing of external and internal 'noise' (accidental features), causing the system to attempt to capture this irrelevant, non-repeating noise. We believe that this view is more readily compatible with the broad range of behavioral peculiarities in ASD.

2.9.2 *Unifying theories of ASD, in the face of its genetic and phenotypic heterogeneity*

Several scholars have lamented the overgrowth of unifying theories on ASD, seeing that they fail to deliver a convincing account for every ASD symptom cluster. Heterogeneity in underlying genetics similarly seems to suggest that there is not one but rather a multitude of deficits underlying the ASD pathology (Happe & Ronald, 2008). Finally, and most importantly, phenotypic variability is notorious in ASD (Rommelse, Geurts, Franke, Buitelaar, & Hartman, 2011). This causes, but may also be caused by, difficulties in diagnosing ASD. Questioning the view of ASD as singular entity even further is the fact that "virtually every symptom characteristic of ASD can be observed in children who do not fit this diagnostic category" (Bishop, 1989). This of course, does not necessarily imply that these symptoms when they appear together in ASD are just the result of the "worst of luck". Still, these observations have led Happe and Ronald (2008) to describe ASD as a fractionable triad, with three independent components (communication problems, social interaction deficits, and repetitive and restricted behaviors and interests) coincidentally co-occurring. Only when the three conspire, subclinical signs become clinical symptoms meriting a diagnosis.

Naturally, we agree with Happe and Ronald (2008) that a better characterization of the subcomponents of ASD is much needed, but an intrinsic coherence of the components may only shine through when the appropriate level of description has been found. As we progress towards more realistic models of the mind-brain, we may be able to formulate more fitting explanations of ASD within these broader models. HIPPEA can be considered a first step in that direction.

Furthermore, there may be more coherence in the ASD symptom clusters than these critical authors assume. For example, while executive functioning and attentional deficits may not be specific to ASD (cf. ADHD), the specific pattern of executive capacities impaired and intact may be distinguishable from other disorders, and may have a privileged relationship with social or emotional symptoms of ASD. It is no doubt a challenge to connect social and communicative symptoms to more basic processing differences, due to di-

vergence in the pathways leading to such high-level dysfunctions and to possible compensatory mechanisms saving these capacities for others. Indeed, a truly developmental account like HIPPEA will predict quite some variability in the unfolding of clinical symptoms depending on interactions with the environment.

Finally, heterogeneity in underlying (epi-)genetic and molecular paths towards the syndrome does not preclude the possibility that one main cognitive mechanism is impaired. There is little reason to expect a one-to-one mapping from cognitive processing to neurobiology. The previous section provided possible ways HIPPEA links up with neurobiological evidence.

2.9.3 *ASD in relation to schizophrenia*

Increasing evidence suggests ASD has common genetic risk factors and neuroanatomical overlap with schizophrenia (Carroll & Owen, 2009; Cheung et al., 2010; Serretti & Fabbri, 2013). Intriguingly, a recent theory of schizophrenia (Adams, Stephan, Brown, & Friston, 2013; Fletcher & Frith, 2009) invoked undue high precision of prediction errors to explain positive symptoms in schizophrenia (hallucinations and delusions). The authors proposed that high precision prediction errors cannot be reduced and are propagated to higher levels, where they induce radical updates of beliefs to somehow make sense of them. Hence, they result in the strange world views and delusions.

Briefly, it seems to us that inflexible, high precision prediction errors are a better fitting explanation for ASD than for schizophrenia. Overprecise prediction errors as a fundamental, indeed developmental, characteristic would be present from very early on in life. Hence, the relatively late onset of schizophrenia needs explaining. In addition, overly high precise prediction errors arguably do not sufficiently explain the specific, improbable and utterly bizarre contents of delusional beliefs (Silverstein, 2013). Other things that may be important to consider are the specific level of origin of the prediction errors (conceptual or action vs perceptual prediction errors, Adams et al., 2013; Fletcher & Frith, 2009) and the subjective confidence level (precision) that top-down beliefs can take on (to explain their fervor).

While the cognitive commonality of schizophrenia and ASD may match their genetic and neuroanatomical overlap, it also highlights a central challenge for predictive coding theories of mental illnesses: if they want to provide more than over-accommodating just-so stories for mental disorders, these theories should be able to give good,

constraining explanations for the cognitive and neural specificities of each disorder. Clearly more work is needed in this respect.

2.10 CONCLUSIONS

While one core deficit is unlikely to explain all heterogeneity in ASD, it is quite remarkable that our approach can accommodate a broad range of reported deficits and peculiarities. This also makes sense since meta-learning is central in development across domains. Metacognition, conceptualized as the ability to monitor and adaptively use uncertainty, is generally fragile, costly and only conclusively demonstrated in a few, cognitively higher developed species (Carruthers, 2008; J. D. Smith, 2009; J. D. Smith, Coutinho, Church, & Beran, 2013). Dysfunction of this capacity may impact higher-level functions such as emotion processing and social cognition, but it also has a pervasive effect on attention, cognitive control, perception and learning. Hence, HIPPEA is broader than earlier single-deficit accounts of ASD, because it is not linked to a certain symptom cluster. At the same time however, HIPPEA is more specific than those accounts, homing in on the disturbed mechanism.

Every existing neurocognitive theory is criticized for not being universal and not being specific for ASD. How does how HIPPEA fare on those accounts? First, does HIPPEA maintain that *every* individual with ASD shows inflexibly high precision of prediction errors (*universality*)? We argue that this is indeed the case, but leave room for two ways to arrive at this high precision: A direct, possibly neuromodulatory deficit in the precision mechanism, or a deficit in the extraction from experience of information that should be used to estimate precision (meta-learning). Second, does every individual with chronically high precision prediction errors suffer from ASD (*specificity*)? Again, we answer positively, but with the important qualification that HIPPEA is consistent with the existence of a spectrum of ASD traits. It distinguishes different perceptual, cognitive, emotional, and social processes according to the extent to which they can be affected by chronically high precision errors. This naturally leads to the notion of a 'spectrum'. Just how high and how fixed precision is, determines whether normal functioning is still possible. Indeed, some people may be able to turn their 'deficit' into an asset in tasks that benefit greatly from their specific processing style (Gonzalez, Martin, Minshew, & Behrmann, 2013).

Evidence-based treatments and psycho-education for ASD that focus on early learning (such as applied behavioral analysis, Lovaas, 1987; Rogers & Vismara, 2008), could take inspiration from HIPPEA that also has learning at its core but demarcates the circumstances

under which problems in ASD arise. Animal models of ASD-related diseases show that environmental enrichment can reduce risk of developmental disorders (G. Dawson, 2008). We also remarked that people with ASD may be able to learn and use high-level predictions, given extensive exposure to more and different situations. However, most of all, our approach reaffirms the importance of more scaffolding during learning (e. g., Bellon, Ogletree, & Harn, 2000; Odom et al., 2003). Our Section 2.3 on exploration made it clear that children with ASD need more support with the gradual progression from simple to naturalistic stimuli (e. g., using virtual environments), taking into account uncertainty *and* its causes. Finally, and slightly counter-intuitively, reducing intense concentration on learning experiences, preventing subjects from trying to match all details ('early stopping'), has also been proposed to be beneficial (Bakouie et al., 2009).

Although we consider HIPPEA a rich and promising theory, much of what we have offered here is post-hoc. The specific theory of ASD we proposed in this paper is based on predictive coding in normal functioning, but so far most of the explanatory power is in our selective (albeit broad) synthesis of the literature on ASD. Future research will have to corroborate its unique predictive power. In the preceding sections, we have often added comments about shortcomings in the current literature as well as specific hypotheses derived from our theory that remain to be tested. With a very general theory like predictive coding, there is always a risk of non-falsifiability (see also the extensive discussion sparked by Clark, 2013b), but we are convinced that our theory of predictive coding in ASD is specific enough to be testable. Although we mainly addressed the functional (psychological) level in this paper, we are optimistic that HIPPEA is at least compatible with an explanation at the neural level. We hope the progress that is currently being made in filling in the neural mechanisms behind predictive coding will help answer the question of precisely why individuals with ASD end up with high, inflexible precision.

In sum, our intent with this paper was to sketch the breadth of implications of HIPPEA with regard to aberrant development, and to point to new empirical questions for ASD research flowing from this view. Ultimately this will give us a better handle on ASD, connecting clinical to neurobiological descriptions and providing a firmer foundation for treatment.

The purpose of art is to force us to notice. Since perception is usually too automatic, art develops a variety of techniques to impede perception or to call attention to itself... Thus, it not only bears meaning, it forces an awareness of its meaning upon the viewer.

— Lemo & Reis, 1965

3

Putting reward in art

It was the Gestalt psychologist Koffka who stated that violations of the law of the good Gestalt "hurt our sense of beauty" (Koffka, 1935, p. 174). If we take 'good Gestalt' (Prägnanz) to mean having regular, clear, symmetrical and simple forms, as it is conventionally defined, this seems to exclude an abundance of artworks (ranging from traditional to modern) from ever being beautiful to human eyes. Why is it that artworks, the prime instances of beauty, often contain precisely these violations which are supposed to "hurt our sense of beauty"?

We will suggest a way out of this conundrum, based on a view on visual art that is firmly rooted in cognitive neuroscience. It will be less reductionistic than other proposals in that art will not be explained by piecemeal activations in separate visual areas on a specific level of visual processing, but by taking into account the bidirectional, hierarchical nature of the visual system as it is implemented in the predictive coding approach. The former approach is doomed to fail for at least two reasons. Firstly, an activation, however strong, of one or more of the visual processing areas is clearly not sufficient for an object or event to pass as a work of art (Hyman, 2010). Melcher and Bacci (2008, p. 357) rightly remark: "A particular artwork may indeed activate area V5/MT, but a passing bicycle would activate those motion-processing areas even more strongly and would not be considered a work of art". Secondly, we are indeed never confronted with the isolated raw visual input, rather our visual system (and the brain as a whole) actively but effortlessly organizes it into surfaces, motions, three-dimensional objects and concepts. Our proposal will center on these processes of perceptual organization (rather than the representational or symbolic meaning of their 'end-products'). The rationale for this is found in Redies (2007, p. 3) who calls for a broadly applicable theory (from cave art to Kandinsky), "as universal as aesthetic judgment itself."

He notes that since almost any (visual) stimulus can be used to compose aesthetically pleasing objects, aesthetic perception may rely on general aspects of information processing that are implemented in all visual channels and regions. Redies adds that aesthetic perception requires the processing of global features encompassing interactions between large amounts of receptive fields. By grounding our proposal in the formal, organizational characteristics of visual information processing as hypothesized by the predictive coding framework, we will attempt to meet these requirements.

Perceptual processes alone would hardly give us aesthetic experiences though. Together with anthropologist Dissanayake we claim that "much is overlooked when aesthetic cognition is conceptualized simply as 'sensory' or 'perceptual'..." (Dissanayake, 2009, p. 163). A one-sided emphasis on perception neglects the emotional aspects that are the motivating drives for creating and enjoying art in the first place. Artistic expressions attract and hold attention and stir and shape emotion. Consequently, any theory of art must explicitly elucidate the crucial interactions between perceptual and emotional processes. Luckily for our purposes, recent data suggests vision may be intrinsically affective (Barrett & Bar, 2009), in the sense that the processing of emotional relevance and value is not an afterthought in the visual processing hierarchy, but can actually drive object formation and recognition. Our challenge will be to specify the emotional consequences of the formal, organizational mechanisms of the predictive coding view, bracketing the emotional or symbolic content of the representations involved.

3.1 PREDICTIVE CODING

The predictive coding approach of perception holds that the brain actively anticipates upcoming sensory input rather than passively registering it. On the basis of prior experience, the brain actively makes predictions about what visual input to expect in the current context of stimulation. At every level of the visual hierarchy predictions are generated and propagated (top-down) to lower levels, where they are checked against incoming (bottom-up) evidence. The idea is that these predictions suppress or explain away the activity in lower areas that agrees with them (de-Wit, Machilsen, & Putzeys, 2010), while what remains and is sent upward are the mismatches between these predictions and the current input, also called the prediction errors. This way the processing resources (attention) can be directed to that part of the stimuli that has not been sufficiently explained (predicted), and thus still has to be learnt. Through constantly fine-tuning predictions using the mismatches, the brain becomes tuned to statistical regularities of our natural visual environ-

ment. These predictions structure the perceptual input in patterns that allow for predictability both within and across visual displays. Thus, the classic concept of Gestalt, traditionally defined as a (experiential) whole that is different from the parts, can be recast in terms of predictive coding (Van de Cruys & Wagemans, 2011a).

This framework is fruitfully adopted to explain several findings on visual processing (but also auditory perception, cf. Kumar et al., 2011; Winkler et al., 2009). It can account for extra-classical receptive field effects measured with single cell recordings in the primary visual cortex (Rao & Ballard, 1999) but also for fMRI patterns of activation across the visual hierarchy. For example, S. O. Murray, Kersten, Olshausen, Schrater, and Woods (2002) observed that perceptual grouping is accompanied by an increase in activity in higher tier object-sensitive areas (LOC) and concomitant decreases of activity in lower visual areas (V1), as predicted by a predictive coding view. Furthermore, the well-known phenomenon of repetition suppression can be explained by a reduction of neural activity for predictable stimuli (Summerfield et al., 2008). Similarly, Alink, Schwiedrzik, Kohler, Singer, and Muckli (2010) cleverly used a apparent motion path to generate strong predictions about when and where a visual stimulus will be, demonstrating that stimuli evoke smaller responses in V1 when they have an onset time that matches these predictions. The more traditional theory of (neural activity in) perception as the piecemeal accumulation of evidence with visual neurons functioning primarily as feature detectors, would have great difficulties in explaining this series of evidence. In contrast, in predictive coding, perception is an iterative matching process of top-down predictions checked against bottom-up evidence along the visual hierarchy. Consequently, each level in the visual cortical hierarchy has a twofold computational role: firstly, it provides predictions (the conditional probability of a stimulus) regarding expected inputs to the next lower level and secondly, it encodes the mismatch between predictions and bottom-up evidence (the prediction error or 'surprise'). Egner et al. (2010) explicitly set out to compare performance of a predictive coding model with the traditional model. They found that when subjects strongly expected to see a face, fMRI activity in the face-selective fusiform area was indistinguishable when actually viewing a face versus a house, while maximally differentiated when the expectation of seeing a face was low. Using computational modeling the authors conclude that this pattern of results can only be accounted for by the predictive coding model, which says that the total neural activity in category-selective areas represents the sum of activity related to prediction ('face expectation') and that related to prediction error ('face surprise').

At the heart of predictive coding is the concern for ease (efficiency) of processing. As described earlier, neural resources needed for processing predictable stimuli are minimized, as our system gradually becomes optimized to the statistics of our natural perceptual environment. In addition to that, the predictive coding strategy is parsimonious because different levels of the processing stream do not need to keep duplicates of information, maintained in higher regions (de-Wit et al., 2010). In light of evolution, being able to encode and process sensory information in an efficient way is vital for the costly organ the brain is.

Being able to successfully predict is another obvious evolutionary advantage, as it allows animals not only to react after the fact to stimuli that change the internal milieu (homeostasis) but also to prepare (anticipate and compensate) for those that are very likely to ensue. Thus, homeostasis urges organisms that can walk around and manipulate their environment to take a predictive stance. Friston's (2010a) generalization of the predictive coding framework starts from homeostasis, or the realization that only a limited set of all the states an organism can be in, is compatible with its continued existence. The long term goal of reducing the time spent in 'surprising' states translates into the short term goal of reducing prediction errors.

This holds for single-celled organisms as well as for complex mammals, like humans, who evidently can rely on a greatly enhanced predictive capacity (Cerra & Bingham, 1998). Based on statistical regularities in the environment, organisms form predictions on the where, when and what of resources in the future, compensating for the inherent variability in the availability of these resources. Complete predictability implies that any disturbances of internal milieu are fully compensated for (by plasticity or action). Propagated prediction errors tell our system whether our current cognitive resources are up to the task of interpreting of and coping with incoming stimuli or whether extra effort is needed. This effort can take the form of learning, when changing the predictions (our internal generative model of our environment) but it can also be behavioural effort when actions are executed to change the things predicted or our sampling of it (Friston, 2010a).

If we acknowledge that predictive coding is ultimately founded on homeostasis and the vital maintenance functions of the body, it should not surprise that we can link this view to emotions. Indeed, to understand aesthetic emotion (appreciation), we need to make this link explicit.

3.2 PREDICTION AND EMOTION

Emotions can be seen as motivational amplifiers (Huron, 2006). They motivate organisms to pursue behaviours that are normally adaptive and to avoid behaviours that are normally maladaptive. When they form accurate predictions, organisms can efficiently react to upcoming events, thereby increasing the likelihood of future positive outcomes. Hence, it would be wise for evolution to reward cases in which predictions are confirmed in actual circumstances. Meanwhile, our failures in predicting situations may be characterized by negative emotion, because they signal that there is something wrong with the mental model we use to generate the predictions. It follows that prediction errors are always to some extent negative in affective valence. In a better-safe-than-sorry strategy, nature tends to assume the worst, because the cost of a false negative (type II error) is potentially much larger than that of a false alarm (type I error). In a conservative reflex, the first, quick reaction to prediction violation or error is negative (Huron, 2006). This is also reflected by the so-called conflict theories of emotion (for a historical review see Mandler, 2003, which claim that emotions arise from interruptions or discrepancies between expected and actual situations. Hebb (1949) and Mandler (2003) are the most well-known advocates of this view.

By studying the continuous intrinsic activity of the brain, Hebb realized that the brain was proactively involved in processing incoming stimuli, rather than just passively responding to them. According to him, thought consists of so-called phase sequences: sequential activations of neural structures (cell assemblies) that are built up as a result of previous experience and learning. Each assembly activation may be aroused by a preceding assembly, by a sensory event, or by both. Negative emotions arise then from the interference with (obstruction of) such an established phase sequence. Although the terms used may seem quite peculiar to modern ears, Hebb's view is consistent with the approach described earlier of implicit formation of predictions and their confirmation or obstruction.

Mandler (2003) elaborated on this theory, arguing that interruptions of ongoing response tendencies and conflicts between expectations and actual circumstances create arousal because they signal important changes in the environment which have to be acted upon. Depending on the cognitive context and the situation the arousal is subsequently evaluated as positive or negative. Two important differences must be noted between Mandler's theory and ours. Firstly, the basis for the generated expectations in Mandler's view are cognitive schemata. We choose the broader term 'predic-

tions' because it can be used for both sensory predictions (in predictive coding) and conceptual (high-level) predictions based on cognitive schemata. Consequently the term prediction is also more neutral with regard to the conscious access to the predictions. Secondly, we described any discrepancy (prediction error) as negatively valenced, while for Mandler discrepancies generate undifferentiated arousal which is later interpreted as either positive or negative. The same discrepancy may produce differently valenced emotions depending on the circumstances and the cognitive context. Instead, we argue that any discrepancy (surprise) is initially experienced as negative, even when a situation is not accurately predicted but in fact better than expected. Instantly afterwards, a reappraisal will take place, converting this into a positive experience and actually more positive than when it would not have been preceded by a negative one. This contrast effect will be made more explicit in our discussions of visual art.

In general, the pleasure linked to visual configurations seems to be dependent on the perceptual and emotional dynamics involved. In a similar view, Kubovy (1999) defines the pleasures of the mind as "collections of emotions distributed over time whose global evaluation depends on the intensity of the peak emotion and favorability of the end". Transitions rather than static states of stimulation are needed for these kind of pleasures. Applied to our context of predictions, positive emotions are experienced when we have succeeded in reinstating predictability (solving the prediction error). In slightly provocative terms we could say that resistance (of prediction errors) breeds liking.

Another way of thinking about this is that gains in efficiency, the sparing use of resources, are rewarded. If the visual system manages to find a sparse explanation of previously unpredictable stimuli this genuinely appears to feel good. Similar thoughts have been expressed in the literature on processing fluency (Reber, Schwarz, & Winkielman, 2004). According to these authors, stimuli are preferred more if they are processed more easily. Because of this, familiar, symmetrical, clear-cut (high-contrast) or prototypical (average) stimuli are thought to be liked most. The mere exposure effect (Zajonc, 1968) is cited as evidence for this position, seeing that improved processing of a stimulus because of repeated presentations leads to increased preference for this stimulus. At least in case of simple stimuli and subliminal exposures, the mere exposure effect has ample empirical support (Bornstein, 1989). Translating this to the terms of the predictive coding framework, an increase in processing fluency amounts to an increase in predictability (reduced prediction error), but the fluency account has largely disregarded

the dynamics of fluency. Recently however, some processing fluency theorists acknowledged the importance of this in the context of the research on the insight experience (the Aha-Erlebnis). Topolinski and Reber (2010) report that a surprising gain in fluency increases positive affect and the judged truth of the solution found. Since this positive affect sets in before any (conscious) assessment of the 'insightful' solution has taken place, they argue this positive effect cannot be solely due to the positive feeling of pride.

Interestingly, discovering the solution of a neutral two-tone image (cf. Gregory's camouflaged dalmatian) is accompanied by amygdala activity and the strength of this activity predicts long-term memory for the organized stimulus (Ludmer, Dudai, & Rubin, 2011). Although the study referred to did not directly probe possible associated (positive) affect, it is well established that the amygdala intervenes in reward processing just as well as in fear processing (Baxter & Murray, 2002; Sander, Grafman, & Zalla, 2003). Here, the amygdala might signal the goodness of the solution found in terms of representational efficiency, and phenomenologically this may be experienced as positive. A reduction of uncertainty, or equivalently, a gain in predictability might account for the positive experience of the Aha-Erlebnis. That perceptual insight, irrespective of the emotionality of stimuli involved as such, is processed by the same neural structures as those responding to the beneficial biological value of stimuli (Baxter & Murray, 2002) is intriguing. It points to the importance of this kind of perceptual and neural (re)organizations in economizing representations.

So far, we infer that a (temporary) state of unpredictability (prediction error) can be as important for the emergence of perceptual pleasure as is the predictability. This might particularly be the case for artworks, in which the presence of discrepancies would otherwise be difficult to explain. Earlier (Van de Cruys & Wagemans, 2011a) we used the example of earworms: the melodies that keep haunting us. They may be great Gestalts (extreme predictability) but they are rarely described as beautiful. We will delve deeper into the role of (un)predictability later on, using artworks as examples.

More neural evidence that (un)predictability is an important factor in emotion has been accumulating. For example, a recent study found sustained activity in the amygdala for temporally unpredictable tones in humans and mice in comparison with rhythmic (predictable) tones (Herry et al., 2007). Moreover, both humans and mice show increased anxiety-related responses in a standard anxiety test when these unpredictable (versus predictable) tones were played in the background. Other subcortical areas also seem to be involved in processing of prediction errors. In research on reward

processing it is found that dopamine neurons in the ventral tegmental area encode the deviations between actual received rewards and the expected reward. The reward prediction errors may be crucial signals for learning about rewards (Schultz et al., 1997). Another region implicated in reward processing, the putamen, was reported to be more active whenever a tone was not followed by the stimulus type it predicted in an experiment only using entirely neutral, unrewarded stimuli (den Ouden et al., 2010). Similarly, plasticity in the amygdala in the context of fear learning, seems to be driven by prediction errors originating in the peri-aquaductal gray (McNally, Johansen, & Blair, 2011). Kapp, Whalen, Supple, and Pascoe (1992) show that amygdala responses wane once contingencies have been fully learned, in other words once stimuli are perfectly predicted, and are only reinstated when external associations change (when predictions fail).

The studies reviewed so far point to an interesting stance, namely that predictions and prediction errors are omnipresent in the brain as has been argued by several scholars (Bar, 2007; Friston, 2010a). We have come across predictions in perception, in reward and fear processing and in cognitive schemata. In social and developmental psychology we encounter a long tradition of theorizing about cognitive schemata giving rise to expectations and about the general tendency to reduce dissonance (see for example Kagan, 2002; Proulx, Heine, & Vohs, 2010), which can also be reconceptualized in terms of predictions and errors. Although a rigorous comparison of these concepts still awaits to be done, we can discern a common theme in all of these studies. It is that the brain's main trade is to predict impending circumstances based on prior, similar experiences. Cases in which this goes awry are significant and therefore emphasized by emotion. The organism has to deal with these mismatches by action (assimilation) or by updating its mental model (accommodation through learning). To show that predictive coding is a pervasive mechanism of mental function might seem of lesser importance for the current purposes. Yet, as an explanation for emotion in art, which evidently works on different perceptual and non-perceptual levels, a theory that encompasses more than purely perceptual principles has substantial appeal.

Finally, this line of reasoning provides a way for visual processing to influence (and be influenced by) emotion. Duncan and Barrett (2007) already remarked that in brain anatomy the boundaries between emotion and cognition seem to dissolve. Conceiving perception in terms of predictions and mismatches opens up a path for perceptual configurations to induce different sequences of affect, partly independent of the particular contents of perception. It is this

link we want to explore with aesthetic appreciation of visual art as a concrete application.

Note that in what follows, we do not aim to explain all possible emotions in art (if indeed such would ever be possible). We focus on perceptual features of art and only briefly touch upon content which can of course induce a whole range of emotions, from very straightforward ones (as for example in paintings of nudes) to very complex ones (as in paintings portraying social settings). Here, we are not interested in intrinsic or associative affective value of the content as such, although we agree that real impressive and expressive art embodies an important interaction between stylistic aspects (perceptual) and thematic ones (see below). Some other affects that we will not talk about but that can show up when contemplating works of art have been listed by Jackendoff and Lerdahl (2006), for example the admiration of craftsmanship, emotions linked to nostalgic memories associated with a particular work, social emotions of belonging to (or differentiating oneself from) a group, etc. These different affects probably interact when viewing art and in the process the sources may be lost, which means that affect caused by one factor can enhance or diminish affect caused by another. Granting that these different emotions are difficult to disentangle in our actual experiences with art, here we are only concerned with the ways in which the perceptual or cognitive organization can provoke emotions.

3.3 AN APPLICATION TO VISUAL ART

Visual art is in many ways different from the visual input we ordinarily receive from our natural environment. We want to focus on the prediction errors, more often called incongruities, ubiquitous in art. When we talk of incongruities in the context of art, surrealistic paintings spring to mind. Or perhaps Duchamps' *La Fontaine*, a porcelain urinal, an object one would never have expected to belong in an art exhibition, let alone to be called 'beautiful'. However, much more subtle kinds of incongruities play in many other styles as well. As a first example, take the painting *Natura Morta* (Figure 4) by Morandi. In his still lifes he often used subtle differences in hues or texture for figure and background, which makes the perception very unstable. The prediction error resides in those dissolving boundaries, and more generally in the fact that Morandi twists expectations of how still-life paintings are defined traditionally.

Why do painters repeatedly and deliberately create such obstacles for us? As we saw, current theories of predictive coding in visual

Figure 4: Natura Morta (1960), Giorgio Morandi

perception assume that our brain aims at reducing prediction errors. In art, however, even though we (or rather, the artists) fully control the stimuli, we intentionally create prediction errors that may not even be possible in the natural visual environment. So in art we do not always manipulate stimuli in ways that reduce prediction errors. Why so? And why does this not imply that we experience most art as unpleasant, as would be the case if all prediction error is negative? The common answer is that in going to a museum, we expect the unexpected. In this limited domain and time span, we can tolerate and even enjoy unpredictability because we expect to be surprised in every new exhibition hall. Furthermore, in these settings there is no need for urgent actions. Prediction errors in art do not need to be acted upon as in real life. This playful and safe as-if-context of art, where our guards can be lowered and our actions suspended, allows for the usually negative prediction errors to be enjoyed. Hence, a positive reappraisal can immediately follow the negative gut reaction. But why take this detour? The frequent use of prediction errors in art and their relation to appreciation deserves to be explicated more thoroughly.

3.3.1 *From auditory to visual pleasures*

We propose that while prediction error is always annoying or unpleasant initially and confirmed predictions are pleasurable as such (mostly independent of their content), prediction errors or delayed prediction confirmation can be an important tool for artists to amplify the subsequent positive affect of prediction confirmation, in a sort of contrast effect (Huron, 2006). For developing the current view we are much indebted to Leonard Meyer and his treatment of emotion in music (Meyer, 1961). He already invoked expectancy violations to explain emotion expressed in music. Tensions are created in music by first establishing a strong pattern in rhythm and melody (a scale based on the tonic) and subsequently deviating from it. The role of expectancy violations might have been realized in music first because musical patterns are spread out in time, so the dynamics of predictions and deviations thereof are easier to notice. However, they also play in visual perception, albeit on a millisecond time scale. For example in *Day and Night* (Figure 5) by Escher, every new bird-like form the visual system encounters, confirms and strengthens the (later destroyed) parsimonious prediction.

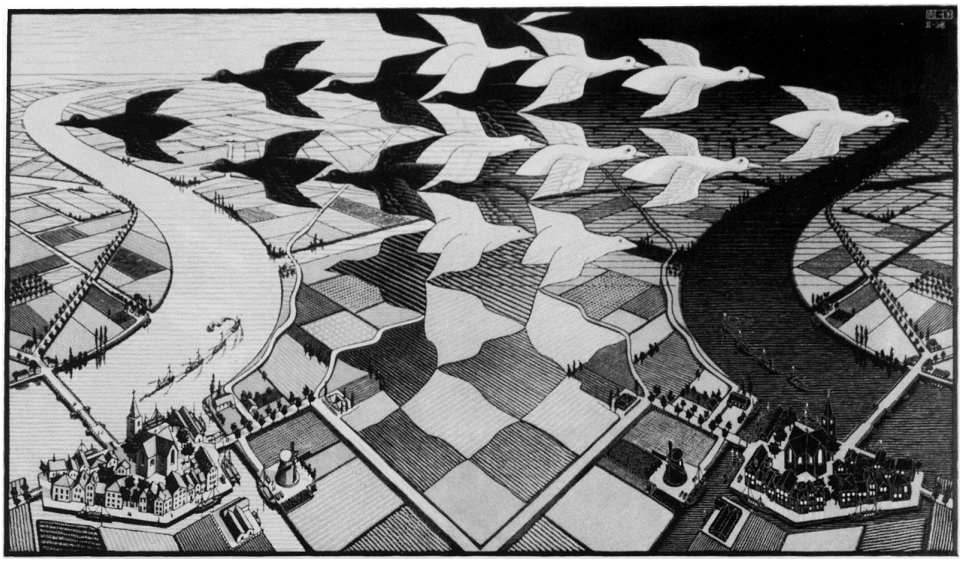

Figure 5: Day and Night (1938), M.C. Escher

An experience of deep aesthetic appreciation is not so easily reproduced in the lab. One more reason music might be more tractable in the lab than visual art, is that music happens to be the art form in which an aesthetic experience can be best elicited in an intense, more or less controlled and reliable (repeatable) fashion. These experiences known as music 'chills' or 'shivers-down-the-spine' have been successfully used to explore the neural and cognitive underpin-

nings of music appreciation (Blood & Zatorre, 2001; Huron, 2006). Cognitively, (Sloboda, 1991) found that these chills are strongly correlated with marked violations of expectation. Neurally, Blood and Zatorre (2001) reported that they are associated by increases in cerebral blood flow in regions involved in reward processing (ventral striatum) and decreases in amygdala and ventromedial prefrontal cortex. Similar to the study by Ludmer et al. (2011) discussed earlier, changes in predictability of stimuli elicit a pattern of activity characteristic of processing biologically relevant, survival-related stimuli.

While musicians use uncertainty on the when (rhythm) and the what (melody, volume) to evoke emotions in their listeners (Huron, 2006), visual artists utilize uncertainty on the where (spatial) and the what (object indeterminacy, see below) with the same intent. Below we will discuss a series of examples, elucidating the workings of this artistic 'tool'.

In *Separation* by Munch, there is a (mild) violation of grouping by similarity of color and form, however we have no problems finding out what the objects are. Artists seem to like using strong predictions, either by building them themselves through repetition as in the Escher painting above, or by using a well-known conceptual domain in their painting and thereby relying on existing strong predictions. In Munch's *Separation* (Figure 6), we find this combination of a familiar pattern (strong predictions) together with a minimal deviation of default expectations, which makes the painting attention-grabbing and memorable. Of course there is more to the paintings we discuss than the aspects we zoom in on. We just want to draw attention to a possible commonality in these pieces, which can appeal to laymen and experts alike.

In his *Weeping Woman* (Figure 7), Picasso arguably counts on our specialized face processing systems (fusiform face area) to project its guesses on what belongs where in a face based only on some fragmentary cues. The viewer quickly runs into incongruencies which presumably generate arousal aimed at reducing the prediction errors. This style-induced arousal could add up to the emotionality of the contents, because despite the 'errors' we can still recognize the emotional expression portrayed. We can contrast this with Ramachandran and Hirstein's (1999) explanation invoking the peak shift phenomenon. In this account, Picasso's face derives its emotionality from being a superstimulus, activating visual face processing systems very strongly because inputs from multiple viewpoints are merged into one image. This explanation for art is refuted by Hyman (2010), who observes that artistic depictions often deviate too far from the norm to be examples of peak shift (which is, as Hyman reviews, a very specific phenomenon in animal behaviour).

Figure 6: Separation (1896), Edvard Munch

In our view it is precisely the incompatibility (prediction error) that causes part of the emotionality in this (and other) paintings.

Predictive coding is an intrinsically hierarchical, multi-level model (Lee & Mumford, 2003) in the sense that implicit predictions are generated and checked on every level in the visual system, from low-level feature-related to mid-level configurational predictions, up to high-level concrete and abstract semantic predictions. Importantly, predictions on every level are partly encapsulated (Jackendoff & Lerdahl, 2006), such that even when we are very familiar with a painting (e. g.*Blanc Seing* by Magritte, Figure 8), and the visual input is in fact not unexpected any more, we are subject to these dynamics. That is, our visual system unconsciously computes its moment-to-moment predictions and errors regardless of the viewer's conscious memory.

3.3.2 *Art movements*

Vincent van Gogh in *The Olive Trees* (Figure 9) plays with perceptual grouping by similarity (parallel waves) breaching the borders of the objects as defined by color and by our top-down knowledge of what the objects are in the scene (trees, fields, sky). In this light, differences between art movements and artists might also be inter-

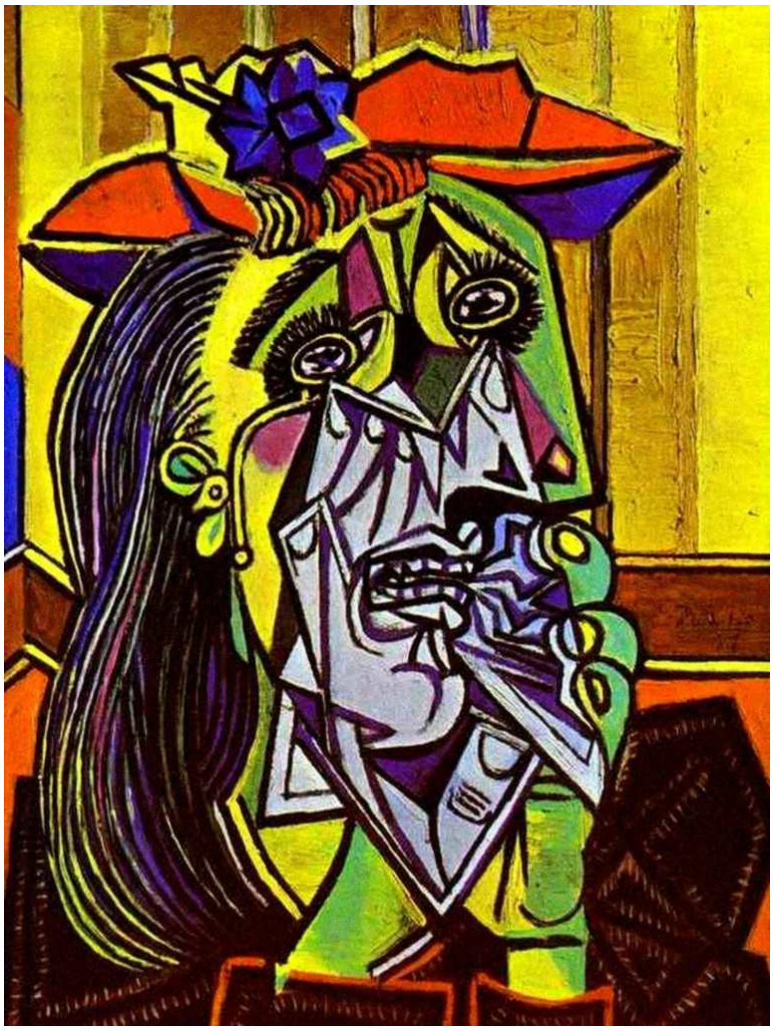

Figure 7: Weeping Woman (1937), Pablo Picasso

preted in terms of differences in the amount and kind of prediction errors primarily used. The evolution of movements can be understood within our framework, recalling that predictions are dependent on the specific history of stimulation, and therefore on cultural and personal experience. Recall that van Gogh (and numerous other great artists) did not enjoy any recognition from his contemporaries. He was mocked and died in obscurity. Predictions evolve and what was a prediction error once can now be the canonical form of expression and thus fully predictable. These shifts partly explain why artistic taste varies widely in different eras and within the same era, between experts and laymen. In the predictive coding approach all perception is a form of expert perception (Clark, 2011), in the sense that it is always determined by an individual's expectations built up through a lifetime of implicit statistical learning.

Figure 8: Blanc Seing (1965), René Magritte

A certain kind or amount of prediction error can be upgraded intentionally or unintentionally to being the norm, the predictable standard, only for those who have developed the allegedly exquisite sensitivities and expertise needed to appreciate or grasp these artworks (ie seeing the underlying predictable structure). This role of aesthetic taste in creating social distinction and status is defended by the sociologist Bourdieu (1984) who argues that by openly declaring oneself an aficionado of difficult and inaccessible artworks people reassert their membership of society's upper classes. It also figures in current theories on the evolutionary role of art as an expensive and time-consuming status symbol, not unlike the peacock's tail (Pinker, 2003).

3.3.3 *Inter-individual differences and the optimum of (un)predictability*

We can derive two important hypotheses from the previous. First, it follows that individuals are likely to have an optimal amount of unpredictability that they most appreciate. Too much prediction error is too unpleasant or even disturbing, none or too little is boring (neither positive nor negative). This relates to psychological or sub-

Figure 9: The Olive Trees (1889), Vincent van Gogh

jective complexity, an important determinant of people's aesthetic evaluations (Gaver & Mandler, 1987). The psychological complexity is a function of both actual stimulus complexity and personal experience with the class of stimuli involved. The predictability of the work of art increases with the observer's knowledge of the correlational structure built up through exposure. An optimum of mild violation of predictions will be experienced as most pleasant, because the experiencer manages to return to a familiar mental schema. Therefore, for a given piece of art, we will also see an evolution of liking through repeated exposure, with initially an increase of preference and motivation to enjoy the piece time and again, but as predictability further increases appreciation will diminish again. A similar evolution of preference has been reported by Berlyne (1970) relating complexity and novelty to arousal. He hypothesized that stimuli of moderate complexity engender a moderate amount of arousal, which is optimally pleasing. Berlyne does not further address the question of whether this mild arousal can be pleasant in itself or, alternatively, whether the moderate complexity still allows a resolution to a predictable, recognized configuration which only then is experienced as positive. We will return to this issue later on.

Reliable inter-individual differences in appreciation should thus be present, which will be dependent on personal experience, which is especially obvious when comparing aesthetic evaluations by experts with those of laymen (Lindell & Mueller, 2011). Additionally, we speculate that these inter-individual differences in preference also depend on the strength of top-down predictions the viewer generates (irrespective of the particular perceptual predictions). This factor may determine how much prediction error a particular person can 'tolerate' and appreciate, but whether this is the case is still an open, empirical question. Related to that, one may wonder whether personality factors such as personal need for structure (Thompson, Naccarato, Parker, & Moskowitz, 2001), field dependency (Witkin & Goodenough, 1981) or the systemizing quotient from autism research (Baron-Cohen, Richler, Bisarya, Gurunathan, & Wheelwright, 2003) relate to this strength of top-down predictions. At least for the personal need for structure construct a correlation has already been found with art appreciation (Landau, Greenberg, Solomon, Pyszczynski, & Martens, 2006).

The second hypothesis derived from our theorizing is that artists attempt to strike this optimal balance between predictability and surprise in their works. This way their viewers have to make an effort and initially experience minor negative affect, only to experience a much intenser positive affect by contrast, once they actually mentally 'resolved' the prediction error. The effort or mental work one has to do to cope with the prediction error is a conditio sine qua non for receiving perceptual pleasure of a Gestalt formation (prediction error reduction). Before arriving at a clear, coherent interpretation of the jumble of lines in Klimt's *Reclining Woman* (Figure 10), our visual system is embroiled in a small struggle. Only with some mental work is the familiar silhouette discovered. But the aesthetic pleasure is larger as a result. Sometimes this Aha-Erlebnis only emerges after a glance at the title for some contextual information, as for Picasso's *Guernica*.

The hypothesis that aesthetic pleasure lies in forcing the brain to do some work has also been put forward by (Dodgson, 2009). His research starts from the computer generated geometric patterns inspired by the early work of the English Op artist Bridget Riley (Figure 11). In a psychophysical test using different degrees of distortion he shows that there is a range in which a pattern is not immediately recognized but can be recognized given some extra effort. Since Riley's works precisely lie in this region where a pattern is hinted at but not made obvious, Dodgson suggests that this might be a common mechanism in all art. In line with our account above, he argues that finding an aesthetic optimum is not a matter of depicting the

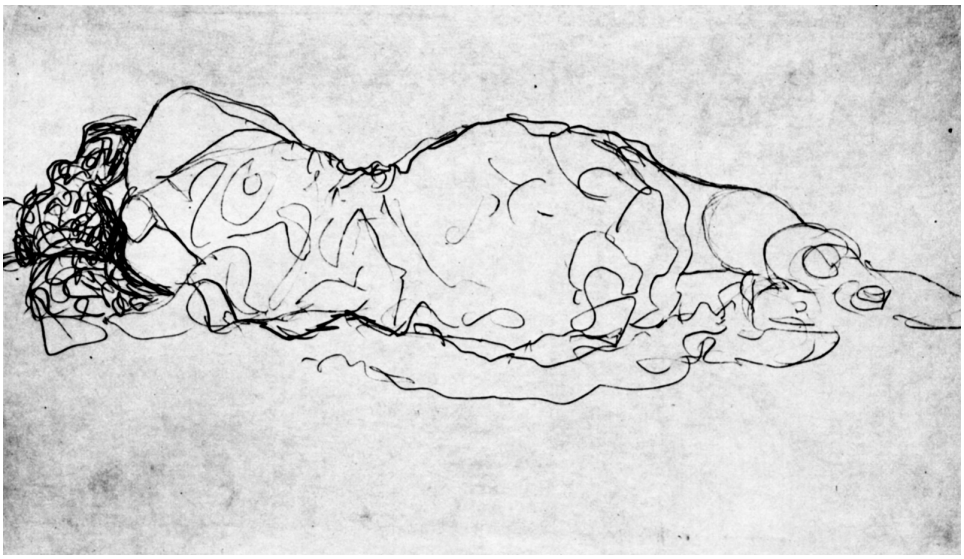

Figure 10: Reclining Woman (1914/17), Gustav Klimt

best Gestalt, but rather of providing just enough information that the viewer can reconstruct the pattern but not so much information that the pattern is plainly obvious. Hence Dodgson's research is a nice illustration of the idea that artworks embody a right balance of the expected and the surprising. It also helps to clarify discussions on the relative importance of familiarity and novelty in art appreciation where it is often concluded that: "It is not extreme novelty but 'optimal' innovation —novelty that allows for the recoverability of the familiar— that is most pleasurable." (Giora (2003, p.176) quoted in Lindell & Mueller, 2011).

3.3.4 *The valence of prediction error*

An central assumption of our theory is that prediction errors are always to some extent emotional, more specifically negative in valence. We are currently setting up experiments to test this, but so far we only have indirect evidence to support it. For example, we note that the expression of surprise, the emotional reaction to violation of predictions, is very close to (and often indistinguishable from) a fearful expression. Neuroimaging studies lend support for this similarity since the amygdala responds equally strongly to surprised as to fearful faces and much less to angry or happy emotions (Whalen, 1998). Moreover, a very recent meta-analysis of neuroimaging studies on aesthetic appreciation (across modalities) found a consistent activation of the anterior insula (S. Brown, Gao, Tisdelle, Eickhoff, & Liotti, 2011). This structure is known to be involved in the conscious, bodily emotional experience (Craig, 2009), in particular for

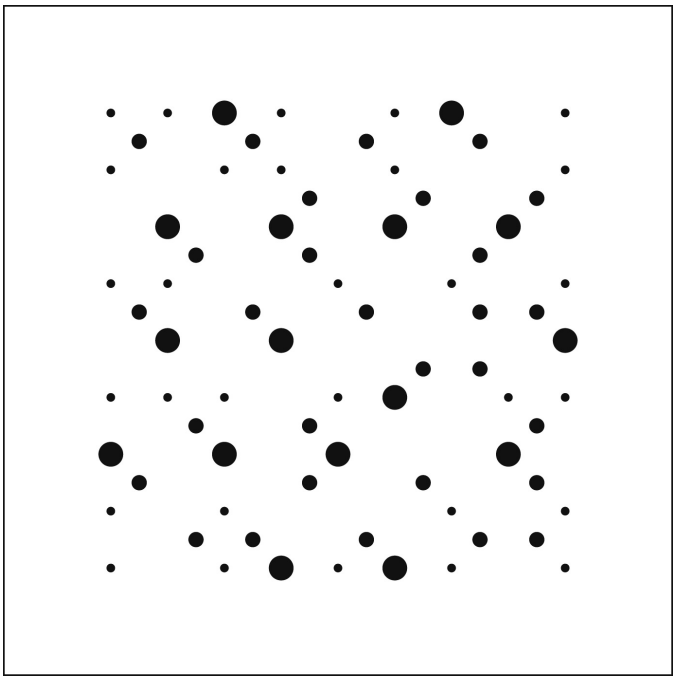

Figure 11: Computer rendition of Fragment 6/9, by Bridget Riley (from Dodgson (2009) with permission from the author)

unpleasant emotions such as disgust and pain. The authors were puzzled that this region systematically pops up in studies on aesthetics, and assumed it attests to the visceral impact of artworks. In the light of our theory, the insula activity might point to the unpleasantness of prediction errors which are an important stage in the aesthetic experience. The limited temporal resolution of fMRI might not allow to make the distinction between stages which quickly follow each other. However, we should tread carefully in interpreting the anterior insula activity, because, as (Yarkoni, Poldrack, Nichols, Van Essen, & Wager, 2011) report in their large-scale meta-analysis, the insula is one of the regions found to be active in a large percentage of neuroimaging studies, a lot of which do not even study emotion. Still, it is interesting to see that the anterior insula is not only involved in negative emotions, but that more specifically (spatially and temporally) separable signals can be found reflecting risk prediction and risk prediction errors in the insula (Preuschoff et al., 2008). This role of the insula in uncertainty and anticipation, which has been noted before (Elliott, Friston, & Dolan, 2000), is in line with our reasoning that prediction errors in art cause uncertainty (risk of not being rewarded in the end) the computations of which will determine whether to look away or to go on exploring the piece. Moreover, the anterior cingulate cortex, a region often jointly active with the insula (Craig, 2009), has been found to be involved in sig-

nalling visual prediction errors (Noyce & Sekuler, 2011) analogous to its role in signalling motor mistakes (which we can reasonably assume are experienced as negative). All this being said, whether our hypothesis of the negatively valenced prediction errors stands the empirical test, remains to be seen.

While traditionally the aesthetic experience is equated with attaining harmony and successful insights, we are not the first to emphasize negative affect, discrepancy and disruption as important components in aesthetic appreciation. Vygotsky (1971) used Aristoteles' concept of catharsis to describe that in aesthetic experiences we go through a transformation from a negatively valenced obstruction to a positively experienced resolution, in full awareness that the positive catharsis would not be possible without first having experienced its contrast. More recently the artist Pepperell (2011) characterized his aesthetic experience as a moment in which his "usual conceptual grip on the world failed." This was accompanied by mild anxiety and "active struggle to make sense of what I was seeing." Artists use the compulsive interpretative (predictive) mode of the visual system to leave out parts of, or blatantly contradict the suggested content of their works.

Pelowski and Akiba (2011) also highlight disruptions and breaches as vehicles for the kinds of self-transformation described in accounts of rich aesthetic experiences. When artworks forcefully challenge the conceptual classifications and personal self-understandings of the perceiver they can lead him or her to question, change or expand these, with possible existential ramifications. This is a precarious balance to strike because the perceiver might prematurely break off perception. Overcoming initial negative experience requires courage and inventiveness to modify the existing (self-)schema (Pelowski & Akiba, 2011; Vygotsky, 1971) but is constitutive for the full aesthetic experience. It is encouraging to see that, starting from different inspiration, namely the full, rich aesthetic experience and several social-psychological findings, the authors arrive at an understanding of the aesthetic experience that is consistent with ours. While their view explores the existential and meta-cognitive implications of the conflicts present in art, ours starts from a substantiated theory of visual perception, in a rather 'bottom-up' way.

3.3.5 *No pain, no gain?*

What evidence do we have to argue that a reduction in unpredictability (prediction error) is experienced as positive? Put differently, do we have reason to assume that the degree of our mental efforts to compensate for this unpredictability (to resolve the picture) is di-

rectly related to the reward we experience? A common finding in animal learning tasks is that random intermittent rewards (there is a pattern but it is disrupted) or non-reward (during initial extinction trials: there was a clear pattern, but now it is interrupted) provoke most vigorous responding. In those cases motivation is highest. For example, a variable interval reinforcement schedule, in which the animal is rewarded on average every n^{th} amount of time, provokes the highest response rates, presumably because from the point of view of the animal, there remains something to be learnt. Neurally, unexpected compared to expected rewards, are associated with increased dopamine peaks. Johnson and Gallagher (2010) specifically asked whether it might be the effort required for getting a reward that is the determining factor for the reward value. They reported that the positive affective quality of taste can be boosted by increasing the amount of effort required to obtain it. In their study mice learned to press a lever A for a reward and lever B for a another, different tasting, reward. Next, they gradually increased the number of operant lever presses needed to obtain one of the two rewards. When tested afterwards outside of the training environment, the mice showed a clear preference for the reinforcer for which they had to work the hardest in the learning phase. In accordance with our speculations on art, they concluded that pleasure (hedonic value) is increased by effort. In humans something similar has very recently been reported as the IKEA effect: participants value the products of their own labour similar to creations of experts, but only when they successfully completed them (Norton, Mochon, & Ariely, 2012).

The view outlined here is bound to learn a lot more from advances in (animal) learning studies on the link between uncertainty versus predictability and reward value. They might help to explain why we can derive pleasures from mildly deviant stimuli, as in an addict who experiences pleasures from almost winning (near-misses). In a set-up mimicking a gambling machine, almost winning is shown to induce heightened reward expectancies in rats, mediated by dopamine and analogous to the near-miss effect in addiction (Winstanley, Cocker, & Rogers, 2011). It seems that part of the reward lies in the anticipation instead of the consumption (Lauwereyns, 2010), with a little unpredictability adding to the reward. One must note that this is in agreement with the idea that biological organisms are characterized by striving: It would have made little sense for evolution to reward static states of stimulation. Prediction errors introduce uncertainty that will stimulate further processing (mental effort), the outcome of which is also uncertain, thus unexpected when successful. Phenomenally, this kind of perceptual

problem solving is not only rewarding but also gives the viewer a sense of mastery (Leder, Belke, Oeberst, & Augustin, 2004).

In the context of art perception, Ishai, Fairhall, and Pepperell (2007) find a positive correlation between the time needed to comprehend a picture and its aesthetic value (measured as the judged degree of 'powerfulness' of a presented painting), leading (Pepperell, 2011) to suggest that the effort people have to invest to recognize the contents has a positive influence on aesthetic value. Reports that the better the viewer's understanding of an artwork, the more he or she experiences pleasure, should thus at least be re-examined (Leder et al., 2004). No doubt individual differences determine a large part of the variability here (as described earlier) but in addition to that, experiment designs may not always allow the tracking of these dynamics and the presence of moments of misconception. It may be true that understanding facilitates liking but this pleasure may be potentiated by preceding mismatches. Likewise, when going to a scary film, we endure cruel tensions because their resolution is so good. Note however, that when we say we like a painting, there is a kind of misattribution taking place: the pleasurable experience associated with a processing characteristic, namely the reduction of prediction error in the stimulus, is ascribed to the stimulus itself (Huron, 2006).

Our fundamental thesis is that the progression from a state of more unpredictability to a state of less uncertainty is pleasurable. Again, animal studies support this hypothesis. In an elegant experiment Bromberg-Martin and Hikosaka (2011) showed that monkeys have a preference for cues that indicate that advance information on the presence or absence of a forthcoming reward will be available over cues that signal such information will not be available, even if their choice has no influence whatsoever on the actual appearance of the reward. Moreover, the monkeys more consistently chose this advance information over the no information cue, than they choose a more probable reward over a less probable reward (Niv & Chan, 2011) In other words, information or reduction of uncertainty is rewarding as such, consistent with our account. In fact, when Bromberg-Martin and Hikosaka (2011) examined the neural underpinnings of the value of information, they concluded that it is encoded by the same habenula and dopaminergic neurons that also encode primary rewards. To be clear, these dopamine and habenula neurons are known to signal a change in reward contingencies (reward prediction errors), and this is also what they found with regard to information: these neurons do not encode the value of predicted information per se but only the changes in predicted information (Niv & Chan, 2011). The signals related to information

change were simply added up to signals related to reward changes, indicating that the value of reward and that of information use a interchangeable neural currency (Niv & Chan, 2011). In keeping with this, we contend that decreases of (information) prediction error are pleasurable and constitute part of the aesthetic pleasure. Cues that signify a reduction of uncertainty also seem to have greater value for humans. Ogawa and Watanabe (2011) asked people to perform a contextual cuing task in which targets are surrounded either by a repeated (and thus predictive) configuration of distractors or, in the other half of the trials, by a novel display of distractors. When subjects had to evaluate the goodness of the displays afterwards, predictive displays were judged more positively than non-predictive or novel configurations. Since people did not see predictive displays more frequently (nor did they recognize them significantly more in a final phase of the experiment), this increased liking could not be caused by mere exposure. The authors conclude that predictability promotes preference (Ogawa & Watanabe, 2011). In the context of an effortful search task, with a large degree of uncertainty on the position of the target, predictable configurations will become associated with a higher value than configurations that do not reduce uncertainty.

To sum up, dealing with unpredictability requires effort from the viewer. But when successful, it leads to positive appreciation. The idea that the viewer's personal efforts and 'accomplishments' matter in art appreciation is also recognized by Mamassian (2008, p. 2152): "not all ambiguities in paintings are re-solved, and artists probably strive to leave the right amount of ambiguities to let the observer contribute to his experience in a personal way." However, only by using minimal prediction errors painters can ensure that viewers will obtain their reward(s) and not give up prematurely. Final gratification may be further postponed as long as the artist has hidden enough 'micro-rewards' in the painting which the viewer can discover along the way.

3.3.6 *Dynamics in art*

Our view emphasizes the role of the dynamics of perceptual processing in art appreciation. This emphasis is not particularly new, but has been foreshadowed by Arnheim although arguably in a less articulate way. In his seminal work, Arnheim (1974) asserts that visual experience is dynamically governed by 'attractions' and 'forces' or 'tensions'. Even though these terms are kept rather vague in his writings, our prediction error account could be seen as an effort to give his concept of 'tension' new substance. Arnheim realized that

these tensions are inherent in any percept as with our concept of prediction (error), even acknowledging the active role of the observer and his past experience in (automatically) generating these tensions. Analogously to our view, Arnheim contends that psychological, as physical, systems "exhibit a very general tendency to change in the direction of the lowest attainable tension level" (Arnheim, 1974, p. 14). A crucial difference between Arnheim's view and ours is that the ultimate goal for Arnheim is balance or stability, while for us it is the reduction of prediction error (or maximization of predictability). In general, once he acknowledged the dynamic nature of perception, he remains overly focused on balance and stability as the static pinnacle of beauty, largely ignoring the emotional dynamics of the transitions going from 'tensed' to 'relaxed' interpretations of visual input that, according to our view, may cause positive appreciations.

Because it is inspired by predictive coding, our view also shares with Arnheim's the concern for parsimony (efficiency of representation, cf. supra), which he in turn inherits from the Gestalt tradition, with its emphasis on simplicity and Prägnanz. Predictability implies more efficient, higher level representations, freeing up expensive processing resources. That there is positive affect (appreciation) associated with this progression has been anticipated by Eysenck (1942, p. 358) in his law of aesthetic appreciation: "The pleasure derived from a percept as such is directly proportional to the decrease of energy capable of doing work in the total nervous system, as compared with the original state of the whole system". Thus, a change towards a more efficient state of the dynamic perceptual networks given the current constraints of stimulation, tantamount to a decrease in prediction error (surprise), is pleasurable as such, as we described for the Klimt sketch above.

3.3.7 *From traditional to modern art*

Melcher and Bacci (2008) draw attention to the fact that artists often use very familiar perceptual domains like faces or animals, but apply their 'artistry' on them. Since prehistoric times, artists exploit very familiar patterns, for which we developed exquisite, specialized systems through evolution and experience. Melcher and Bacci (2008, p. 352) observe that "the ability to recognize certain stimuli quickly and easily makes it easier, then, for artists to add decorative elements, accurate details and artistic style." For example in prehistoric cave art (Figure 12), the patterns portrayed are not 'just' depictions of animals or humans, but they incorporate counter-intuitive or unusual deformations (see examples).

Figure 12: Examples of Prehistoric cave art (elongated human figures in aboriginal art in Kukadu National Park, Australia)

The cultural anthropologist Dissanayake (2009) calls this special treatment *artification* or 'making special' which, according to her, encompasses formalization, exaggeration, elaboration, repetition, and manipulation of expectation. Note that most (if not all) of these terms relate to either confirmation and strengthening of predictions or violation and disruption of predictions (prediction errors). Dissanayake agrees that this process is most obvious in the temporal, performing and traditional arts, but claims that these are universal characteristics common to all art forms. This particular shaping of stimuli is meant to set these objects or practices apart from and make them more than the ordinary (traditionally in the context of religious rituals).

In the context of their discussion of surprise in art, Melcher and Bacci (2008) provide some other fine examples of prediction error in art. About the *Sleeping Hermaphrodite* they write: "Displayed with the back to the viewer, it offers a sensual pose that flatters the femininity of the figure, in accordance with the tradition of the reclining Hellenistic nude. Enticed by the sensual nature of this work, the viewer who will then walk around the sculpture to see it from the front will encounter the surprise of the figure's androgynous nature, which provides a moment of astonishment in which sensory information does not coincide with expectation." (Melcher & Bacci, 2008,

p. 354). On the powerful, realistic evocation of biological motion in Caravaggio's *The Death of the Virgin* they hypothesize: "Thus, perception of motion in static art involves a form of perceptual ambiguity, in which complex motion areas such as MT/MST are activated but 'early' vision detectors in V1 are silent" (Melcher & Bacci, 2008, p. 357).

Several theorists and art critics have argued that (post-)modern artists, with their explicit abandonment of any representational aspirations and their blunt refusal to provide any familiar reference point whatsoever to their spectators, have taken things too far (Landau et al., 2006; Pinker, 2003). They are said to leave the viewers of their works completely in chaos, bombarding them with colours and fragmented shapes with no structure or meaning. In our terms: an overload of prediction errors, that despite our best efforts, cannot be reduced to a more predictable, sparser explanation. As a consequence modern abstract art, as compared to other styles, is least preferred by the general public (Landau et al., 2006; Lindell & Mueller, 2011). We suggested earlier that artists will attempt to find an optimum, but their success in doing so depends on their intended audience and its broader cultural context. Works of art will survive through time if they manage to be at such an optimum (and thereby appeal) across cultures and periods (Melcher & Bacci, 2008). Of all abstract art our theory predicts that we will be most impressed by paintings in which we can, with some labour, distinguish recognizable, predictable forms. *Excavation* (Figure 13) by de Koonig shown below can serve as a good example, as can the paintings of Pepperell that also show this kind of object indeterminacy, investigated by Fairhall and Ishai (2008).

Clearly our approach can be applied to art belonging to a broad range of styles and eras. In fact, it has most difficulties in explaining the attraction of hyper-realistic art. In this case we can speculate that it has more to do with the admiration of craftsmanship, the emotional content of the painting, or possibly an ultimate prediction error in all visual artworks: this work provokes strong feelings in me, as if the particular object depicted is really in front of me, while in fact it isn't. What we could call the '*ceci n'est pas une pipe*' experience. Alternatively, we might be bemused by the fact that we see a nearly photo-realistic image while we expected a painting. But evidently these are just ad hoc explanations.

3.3.8 *The fate of prediction errors*

The examples so far might have made it clear that often we do not end up with a coherent, predictable Gestalt. Still, we want to draw

Figure 13: Excavation (1950), Willem de Koonig

attention to the fact that prediction errors at the level of style (perceptual ones), sometimes can be resolved on the level of meaning (see also Pinna, 2010). The deserting lady in Munch's *Separation* dissolves in the road and the air and Picasso's Weeping Woman can be 'broken' because she is sad. A clear congruency in meaning unexpectedly saves the coherence of the paintings and so the prediction error is reinterpreted. With these symbolic explanations we enter a more speculative realm, but it would be foolish to expect that something as complex as art appreciation can be understood by only looking at the low-level, perceptual features. Art is about the interaction of style and content, not a simple addition of these components. As we reviewed in the introduction, a predictive coding view can, similar to schema theory (Proulx et al., 2010), provide a way for high-level expectations to be involved in our experience with art.

What about those paintings in which we can not even remotely and mentally come to a form of closure? Paintings are static art forms, so prediction errors often cannot be resolved, except in our minds. In their artistic endeavours people seem to deliberately seek prediction errors. Prediction errors intrigue us, especially when they violate strong default expectations. Even if we are not sure there is an actual clue to them, we cannot remain indifferent to them (they

cause arousal) and we keep coming back to examine them (attentional resources are recruited). We might still experience positive emotions from these 'unsolvable' paintings though. By reappraisal of the negative prediction error in a safe context, the resulting emotion is still very positive because of the contrast effect (cf supra). A related, but more implicit mechanism, namely misattribution, could turn a negative arousal into something of positive affective valence. This mechanism shows people often have very little insight in the sources of their experienced emotions or arousal, as exemplified by the classic Capilano suspension bridge experiment (Dutton & Aron, 1974). In this experiment passers-by on a scary rope bridge or a solid wood bridge are asked some sham questions by an attractive female researcher and told that they can always call for more information on the research afterwards. The authors found that if participants were questioned on the foot-bridge, they were much more likely to call the researcher afterwards to ask her out. The authors assumed that those participants misinterpreted their arousal from the fright of walking on a high, shaky bridge as feelings of attraction. The specific experiment has been criticized but the phenomenon of misattribution has been replicated since and is well-established. It might thus partly explain why we misinterpret for example the mildly negative color mismatch in a Matisse painting (Figure 14) as affectively positive.

The question of whether prediction errors can be experienced as positive also relates to the conceptual confusion on what exactly an aesthetic experience entails. We can safely assume that art has to be rewarding in one way or another, otherwise we would not be motivated to engage with it. But people might label some artworks as 'fascinating' or 'special', rather than 'beautiful' (Augustin, Carbon, & Wagemans, 2012). Finally, a particular painting might also be differently appreciated depending on the context of stimulation, as we will briefly discuss in the next section.

3.3.9 *Inferentially rich, attention-grabbing meaning threats*

Because of the prediction errors, we feel impelled to question our perception and to linger on its contents. These visual or cognitive challenges urge us to, implicitly or explicitly, go through multiple cycles, exploring different predictions and the corresponding errors (Leder et al., 2004). They grant acces to different layers of meaning, which we so much like to discover. They create the multi-interpretability and ambiguity that has been invoked by others to explain our enjoyment of art (Biederman & Vessel, 2006; Mamassian, 2008; van Leeuwen, 2007; Zeki, 2004). In the context of cre-

Figure 14: Blue Nude (1952), Henri Matisse

ative discovery, Verstijnen, van Leeuwen, Goldschmidt, Hamel, and Hennessey (1998) have coined the term *surplus structure* to denote that through externalizing their ideas in sketches, artists themselves discover new, unanticipated features and interpretations of the raw ideas. Might this be true, not so much because in sketches they can more lucidly represent their ideas, but rather because sketches allow them to depart more easily from ordinary ideas and exaggerate, restructure and deform more freely, in other words amplify what would normally be prediction errors?

Biederman and Vessel's (2006) idea is basically that inferentially rich stimuli will be preferred because they are accompanied by more activity in regions higher up in the ventral visual stream, which possess higher amounts of mu-opioid receptors. This hypothesis is not at all incompatible with our account. Our optimally deviant expected patterns with a limited violation of intuitions, combine the rich inferences (of a predictable pattern) with the high saliency (of the discrepancies) which conspire to make a very emotional and memorable stimulus (Sperber & Hirschfeld, 2004). Discrepancies are attention-grabbing and stimulate further processing, but only

when strong predictions are first built up (clear organization). This optimum makes for a highly relevant stimulus according to Sperber's (2005) theory because it guarantees the richest cognitive inferences for the least cognitive effort.

Our account also bears a clear resemblance to the Freudian notion of the uncanny (das Unheimliche). Freud observed that we are disturbed and aroused by unfamiliar experiences in an otherwise completely familiar setting (Proulx et al., 2010). For example, in absurdist or surrealist art we often find an unfamiliar juxtaposition of very familiar objects. Thus, strong expectations have to be present before such an experience can ensue. In a completely unfamiliar situation no strong predictions are formed so no violations will be encountered. Incongruency or expectancy violation will result in negatively valenced arousal aimed at reducing the inconsistency if at all possible, as in our account. Some authors have even reported that this arousal can lead to an affirmation of any other meaning framework to which one is committed. For example, after exposure to an absurd Kafka parable, subjects more strongly affirmed their cultural identity than after reading one of Aesop's meaningful parables (Proulx et al., 2010). The meaning maintenance model, as this theory is called, has been developed to counter the terror management theory, which assumes that only mortality threat (salience) will cause people to defend their cultural world-view. In a study by Landau et al. (2006) terror management theory has also been applied to art appreciation. Apparently, mortality salience (having people to imagine what will physically happen when they die and which emotions they have with that) decreases their liking for 'meaningless', abstract art, while leaving their appreciation of representational paintings untouched. In further studies the authors find that this effect is limited to individuals with a high personal need for structure, and diminished when the abstract artworks were given meaning (e. g.by giving a title). This research goes to show that next to stable traits of the viewer and stable characteristics of the piece of art, aesthetic appreciation can also be influenced by the context-dependent cognitive and emotional mindset the viewer is in. We could speculate that any contextual uncertainty (prediction error or, in terms of the theories discussed, threat of meaninglessness) could add up to the uncertainty in a painting and thus influence its appreciation. For instance, against a background of unpredictable (vs rhythmic, cf supra) tones a representational painting may be liked more, while an abstract painting might be liked less. Related to this, Mueller et al (in press) recently probed implicit and explicit attitudes towards creativity after inducing a sense of uncertainty in half of their volunteers. These people were

told they might receive additional payment based on a random lottery or, in a second experiment, they were primed to be intolerant of ambiguity. While explicit attitudes towards creativity were similar in the experimental and the controls, people in the high uncertainty condition had an unconscious bias against creativity, and judged a highly creative idea less favorably.

Ultimately, maintaining (or returning to) predictability is about survival and maintaining the body through homeostasis (Cerra & Bingham, 1998; Friston, 2010a; Van de Cruys & Wagemans, 2011a). Predictive coding is about reinstating predictability and therefore about affirming one's own existence. So within a predictive coding framework, we do not need to assign a special status to the existential threat of mortality. We further assumed that making progress in this prediction project of life genuinely feels good or, in the case of art, is beautiful. In his Ethics, Spinoza (2001, p. 40) writes "[I]f the motion by which the nerves are affected by means of objects represented to the eye conduces to well-being, the objects by which it is caused are called beautiful; while those exciting a contrary motion are called deformed." When a successful, sublime aesthetic experience is described as a selfless state of harmony between the viewer and the world (Pelowski & Akiba, 2011) we might take this quite literally: the beholder has advanced in tuning the self to the world.

3.4 REMAINING QUESTIONS AND CONCLUDING REMARKS

We hope to have shown that the predictive coding approach can summarize and throw a new light on existing concepts in the flourishing field on aesthetic appreciation, such as familiarity, complexity, novelty, prototypicality, interpretability, fluency, incongruency, ambiguity and so. Further research will have to make clear what the added value is of thinking in terms of predictions and prediction errors in comparison with these concepts. We want to end our overview by discussing some of the limitations and advantages of our approach.

First of all, we do not want to reduce aesthetic experience to the formal mechanisms discussed. For instance, we are not saying we experience a full-blown aesthetic reaction when discovering the actual content or organization of 2-tone images, but our hypothesis is that a significant part of the aesthetic appreciation comes from discovering the organization after struggle.

Second, we are being too vague when we claim that art is about optimally or minimally unpredictable stimuli. The amount but also the kind of prediction errors seems important. Some prediction errors are more potent than others. For example we saw that artists

either induce strong predictions themselves in their viewer, or rely on strong existing predictions of the domain used in the painting, to subsequently violate them. Also, different artists may have different preferences for the kind of prediction errors they use, as reflected in their style. For instance, some artists play with classical grouping principles and their competition.

Third, one might object that our focus on prediction errors is born out of cultural myopia. In Western art there is a strong impetus to be original and novel, and even to defy established traditions. In traditional, non-western cultures however, originality is often discouraged and artists are expected to closely follow and endlessly repeat the same set of patterns, passed on for centuries (Dissanayake, 2008). As we mentioned earlier, this repetitive art also involves a modulation of predictability but it seems to lack any prediction errors. However, Dissanayake also notes that these forms of art originally take place in the context of ritual ceremonies in times of transition or uncertainty. These rituals nearly always concern biologically important things, such as "assuring or restoring subsistence, safety, fecundity, health, prosperity, and victory or successfully dealing with the bodily changes and emotional and social concomitants of sexual maturity, pregnancy, birth, and death." (Dissanayake, 2008, p.19). Hence, she sees stress reduction or coping with uncertainty as an important adaptive function of art. If we assume that, at least in these particular ceremonial situations, people in traditional societies experienced more life-threatening uncertainties than we do in our modern, Western society, our hypothesis of a general preferred optimum of unpredictability could still hold. In traditional communities, art could primarily function as a vehicle for re-establishing predictability. In Western culture, on the other hand, we artificially create obstacles for predictability (in art, disaster movies etc) to be able to experience the joys of their resolution (while we often still use predictable patterns for our wallpapers and decorations). The proximate cause for making and consuming art would then be emotion regulation, not 'just' uncertainty reduction. Also according to Dissanayake, art can be traced back to the simplified, repeated and stereotyped interactions between adults and children, which assist the development of emotional self-regulation, attention and learning. Indeed, the dynamics of prediction and emotion seem to be protracted in children, where, for instance in the in the peek-a-boo game, the contrast effect (positive emotion following a negative one is more intense) is easily observed.

While art may therefore be a form of training of our exploratory learning capacities in a safe, playful context —information foraging is after all a vital human capacity (Vessel, 2004)— it didn't neces-

sarily evolve for that reason. Rather our approach connects to the neural recycling hypothesis (Dehaene & Cohen, 2007), which assumes that art (similar to for example writing) didn't evolve for any particular adaptive function, but is the result of cultural inventions exploiting evolutionarily older brain circuits and inheriting many of their structural constraints. Indeed, there is no art module in the brain that needed to evolve. Artistic abilities are piggy-backing on our perceptual and emotional information processing capacities. Once in place, art may or may not have become a criterion by itself for selective forces to work on (co-optation), for example in mate selection (Pinker, 2003) or as a way to promote belonging to a social group.

Fourth, unpredictability and its resolution are important in other human activities. For instance, games are most rewarding when they have just the right amount of difficulty, of unpredictability. Similarly in humour we build up expectations and create discrepancies (Hurley, Dennett, & Adams, 2011). Even in science we are most astonished when a scholar discovers (and manages to explain or make predictable) a counter-intuitive discrepancy in a very familiar domain. One might wonder what, if anything, is special about art. But does art need such an essence? Could we not suffice by stating that it is a human activity involving the full emotional and cognitive abilities of human beings but with no immediate biological purpose? A quasi-necessary result of a greatly expanded predictive capacity and an extended ability to delay gratification?

Finally, how do we explain within the predictive coding framework that humans —while ultimately aimed at maximizing predictability, or equivalently: minimizing prediction errors— still explore unpredictable stimuli, and even intentionally create them as in the case of art. This is a matter of current debate (Fiorillo, 2010; Friston, 2010a, 2010b). The immediate motivation of seeking prediction errors may, in our view, be obtaining a larger reward (by contrast) later. Friston explains that this complicated exploratory (itinerant) behaviour does not violate the general tenet of minimization of surprise, provided that the agent "revisits a small set of states, called a global random attractor that are compatible with survival" (Friston, 2010a, p.2). His generalization of predictive coding in the free energy principle optimizes this motion through sensory state-space.

Turning to the advantages of our approach, artists and art critics will approve of the importance of the active, albeit largely implicit, role of the subject (the viewer) in the predictive coding approach. Even a static painting becomes a dynamic experience, as (Dewey, 2005, p. 222) writes: "The product of art — temple, painting, statue, poem — is not the work of art. The work takes place when a human

being cooperates with the product so that the outcome is an experience that is enjoyed because of its liberating and ordered properties." Our proposal also repeats another adage of artists: learn the rules well, so you can break them effectively.

Thus, predictive coding seems to agree with how artists themselves think about what they do. The painter Henri Matisse famously said when interviewed: "C'est une création par les rapports: Je ne peins pas les choses, je ne peins que des différences entre les choses." (Aragon, 1971, p. 140). We may not read too much into it if we say that he intuitively appreciated that our brain works contextually, by figuring out the differences (prediction errors), not the absolute values (Nikolić, 2010; Ramachandran & Hirstein, 1999). Any outline sketch is a sketch of the differences that the visual system would pick up when viewing the real scene. In fact scene category can be decoded from fMRI activity in the visual system during the viewing of line drawings of scenes, just as well as from brain activity while viewing colour photographs of the scenes (Walther, Chai, Caddigan, Beck, & Fei-Fei, 2011).

Second, honoring Occam's razor, we may not need a special, separate psychological theory for aesthetics by using predictive coding. This is consistent with the idea that general-purpose motivations and capacities are involved in art, even though their particular combination might be special to art. In this case, we can rely on the cognitive and neural evidence for predictive coding, which is broader than visual perception (Friston, 2010b; Winkler et al., 2009). The latter is particularly appealing because art obviously consists of more than visual perception alone. Predicting is the default-mode of the brain, encompassing perceptual and semantic levels. And even though the emotional implications of the predictive coding approach have not been thoroughly explored, its potential to connect perception, learning and emotion, may be clear from this proposal.

Lastly, our view may open new avenues for the empirical study of aesthetic appreciation. Prediction error and confirmation may be tractable in the lab, and thus may allow us to isolate one mechanism involved in aesthetic appreciation. For instance, we can induce strong short-term predictions in subjects and subsequently violate or confirm them. Also, it may help to have physiological markers of prediction violation, for example in the event-related brain potential visual mismatch negativity (Kimura, Schröger, & Czigler, 2011). Here we can expect that temporal aspects and expertise will critically influence the outcomes in perception and emotion.

In this paper we have outlined a theory of art starting from the hierarchical, bidirectional dynamics of vision. We concluded that it

is not the most predictable stimulus that is most pleasurable (the easiest Gestalt formation, cf perceptual fluency) but the Gestalt that appears unexpectedly after a fair amount of 'obstinate obstruction'. The positive affective evaluations result from a transition rather than a certain state of stimulation. But because our cognitive system ultimately aims to return to predictability, an optimal amount of prediction error exists. Eventually understanding art implies fully understanding our brain (not just the visual system) and its embodied embeddedness in the natural, social and cultural environment. We are far from such an understanding so any theorizing on art is necessarily preliminary and speculative. Our theory on art is really a theory about perception and emotion and their interplay because we believe that only by understanding this interaction we will come to comprehend human artistic behaviour.

The pleasure derived from a percept as such is directly proportional to the decrease of energy capable of doing work in the total nervous system, as compared with the original state of the whole system.

— H. Eysenck, 1942

4

Is Gestalt discovery intrinsically pleasurable?

People love to find structure everywhere. The success of mundane games like Tetris relies on this, but artists also exploit it. Just think of the sketches by Klimt or Picasso that, with a few well-placed lines, evoke vivid subjects or complete scenes (Koenderink, van Doorn, & Wagemans, 2012; Van de Cruys & Wagemans, 2011b). Or consider ambiguous figures like Figure 15. However, once this predictable structure is found (even if it is only imagined), we quickly lose interest. Yes, when explicitly asked, we will often report we like the predictable, well-structured or symmetric stimulus more, but that is presumably only because we generally confuse end-product with the process followed to get there. As we sorely know, it is very hard to repeat positive experiences just by revisiting the now familiar, predictable end-structure. The original pleasure quickly fades away, because the situation has become fully predictable.

This volatile character of (perceptual) pleasures arguably explains the mixed evidence on liking of very familiar, regular or prototypical stimuli (e. g., Bornstein, 1989). These stimuli are quickly considered boring (Bornstein, Kale, & Cornell, 1990) and more importantly, highly attractive stimuli rarely conform to the principle (cf. art). Then Berlyne's (1970) notion of an optimum level of complexity appeals more, and has also been supported by some evidence (e. g., Krupinski & Locher, 1988). However, this idea that there is a moderate amount of novelty, unpredictability or complexity that is most pleasing is still descriptive, similar to the related law of Yerkes and Dodson (1908) on optimal arousal. This line of thinking stagnated because it did not provide many clues as to the why and how of people's appreciations for intermediate levels of complexity. Berlyne and his followers still interpreted a pleasurable situation or stimulus as too static a thing.

Such static accounts were presumably inherited from the Gestalt psychologists, who of course were the first to stress the salient and positive qualities of the 'goodness of form' or *Prägnanz*, i. e., orderly,

balanced and coherent configuration (Wagemans, Elder, et al., 2012; Wertheimer, 1923). In their writings there was, however, a clear willingness to look at the dynamics of Gestalt formation (and self-organization; Wagemans, Feldman, et al., 2012). A lack of adequate theories incorporating dynamics stifled their progress, and experimental work similarly remained focused on static stimuli. In psychoaesthetics, the Gestaltist Rudolf Arnheim made an influential contribution with his idea of 'forces' in artworks, which was meant to be intrinsically dynamic, as tendencies towards a balance (Arnheim, 1974). However, his account remained mired in vagueness. In fact, he ackowledged that: "At which stages of this complex process the physiological counterpart of our perceptual forces originates, and by what particular mechanisms it comes about, is beyond our present knowledge." (Arnheim, 1974, p. 17). He further only referred to Köhler's (1940) discredited field theory to explain the 'pushes' and 'pulls' in visual patterns.

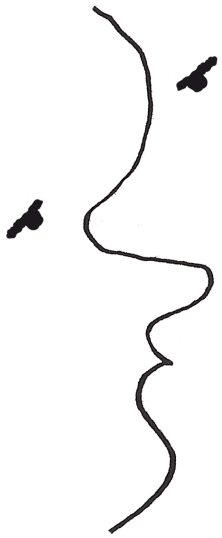

Figure 15: "Two-face" from Hebb (1949)

Nevertheless, these authors put the issue of possible pleasure attached to Gestalt discovery on the agenda. Indeed, 'discovery' entails an active, dynamic contribution of the viewer. Current increasingly popular theories of the predictive mind (Clark, 2013b; Friston, 2010b; Hohwy, 2013) allow for the dynamic reinterpretation that was foreshadowed by Gestaltists. Given the constant, intrinsic predictive activity of the brain at the core of these theories, violations of these predictions (called 'prediction errors') can take on the role of what was previously, rather metaphorically described as tensions or forces. Indeed, these 'predictive coding' theories assume

that prediction error minimization is the main guiding principle for perception (and beyond), where a balance or equilibrium can be attained if top-down predictions match the input, at every level of the perceptual hierarchy (Friston, 2010b). Based on these developments, we proposed an account of perceptual pleasures that focuses on the processing dynamics (Van de Cruys & Wagemans, 2011b).

In a nutshell, we argue that (initial) prediction errors are essential to the pleasure, because the latter depends on predictive progress. In other words, a change in predictions errors (i. e., the first derivative over time) from higher to lower. These transitions are very clear in music where predictable patterns are established, deliberately broken, and returned to (Huron, 2006). If one considers the iterative predictive matching and settling, one can understand that the processing of static (e.g., visual) is never really static, especially since it is often supported by saccades-as-hypotheses (Friston, Adams, Perrinet, & Breakspear, 2012). Prediction errors emerge and are reduced in time. This can happen very quickly, but it can also take quite some time before the right predictions (reinterpretations) are found to explain away the errors. The key thing is that active reduction of prediction errors is required for a pleasurable experience. A common way to do this, seems to be by first increasing the prediction errors, such that larger predictive progress (error reduction) can be made. Or to put it in simple Tetris terms, if there were no annoying protruding blocks, there would not be any pleasure to the 'evening out' with fitting blocks.

Basically, our hypothesis requires three stages, although, as mentioned, these may succeed each other very quickly. In the first stage a regularity is established to induce the corresponding prediction(s) in the observer. If a domain is used for which strong, default predictions already exist in a given observer, this step could be skipped. The second step is the violation(s) of the prediction(s). These can be substantial but the structure has to be recoverable (Dodgson, 2009) for a given observer. The third stage is the 'repair' or reestablishment of the prediction(s), but note that this new pattern is not necessarily the same as the original pattern. Take for example this sequence of stimulus presentations: ABC-ABC-ABC-AB*A*-B-AB-AB. After the prediction error (italic *A*), the participant has to repartition or reorganize the inputs (here: groups of two instead of three stimuli). The 'violation' turned out just to introduce a new predictability.

Two-tone images (also called Mooney images; Mooney, 1957, for an example see Figure 16) are often used to investigate perceptual (re)organization and perceptual 'insight' (*Aha-Erlebnis*). In their best instances they are a form of one-shot learning (cf. Richard Gregory's Dalmatian): once you solved it (or it has been solved for you) you

can never go back to the phenomenal sense of the 'disorganized state'. If one has fitted the pieces together once, one immediately perceives the correct interpretation on the next confrontation with the degraded image. This phenomenal switch forms some of the strongest proof for a substantial top-down, constructive influence in perception (Gorlin et al., 2012; Hsieh, Vul, & Kanwisher, 2010). Prior predictions, learned through exposure with the solution, will define the percept, reinterpreting the black and white blobs that are given. Hence, two-tone images make up strong examples of phenomenal Gestalts, in the sense of that they highlight the contribution of the perceiver in going beyond actual sensory inputs. Interestingly, because traditional Gestaltists maintained that past experience has negligible influence on perceptual organization (Wagemans, Elder, et al., 2012), they would probably downplay the importance of these particular examples.

For our purposes too, two-tone images are interesting stimuli. Two-tones clearly have plenty of suggestion of meaning, so we can assume they sufficiently spur automatic predictive activity in the viewer's visual system. Hence, when given the time, people will come up with all kinds of (often incorrect) candidate solutions. However, any temporary solution considered by the system, will generally have substantial prediction errors. Of course, once the actual gray-scale solution has been given (or sometimes found independently), all these inconsistencies dissolve: the prediction errors are explained away by the activation of the right prediction. Clearly, observers progress from a state of higher prediction errors to a state quasi-complete predictability: all inputs 'fit'. This transition should, according to our account above, be accompanied by positive affect.

Anyone who has experienced the 'insight' moment of a two-tone image has some introspective evidence that this indeed feels pleasurable, but only recently has the question been subjected to empirical research (Chetverikov & Filippova, 2014; Muth & Carbon, 2013). Both studies found that correct identification of the two-tone image was associated with increased liking of these images. There were two methodological commonalities that are worth highlighting to distinguish them from our own approach. Both relied on independent solving of the two-tone images, and the solutions were never shown (nor was other feedback provided). In both these experiments, solving the two-tone was the explicit task for the participants. They also had to explicitly report through rating scales how much they liked the two-tone images, either immediately after the identification question (Chetverikov & Filippova, 2014) or in a subsequent, separate 'appreciation' block (Muth & Carbon, 2013).

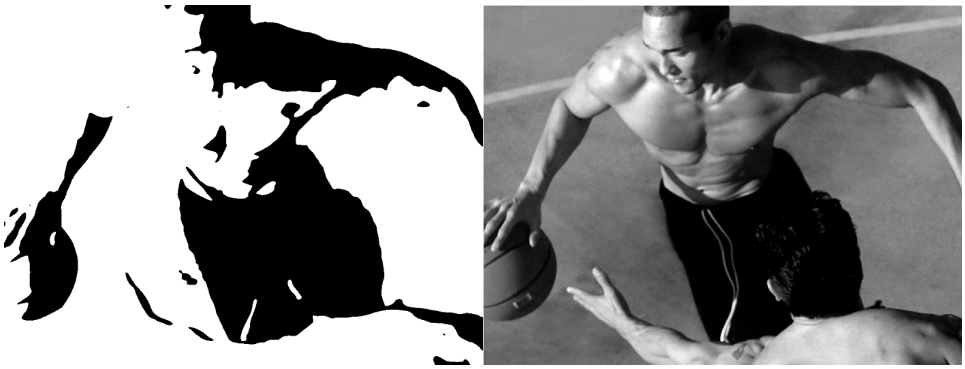

Figure 16: An example of a two-tone image and the grayscale image it is based on.

To avoid demand effects, we chose not to explicitly ask people about their appreciation of stimuli. Therefore, we modified an indirect measure of affect, known as affective priming (reviewed in Fazio, 2001). Affective priming studies find that activating an affective state by showing people positively or negatively valenced words or pictures, called primes, will lead to faster reaction times to stimulus materials (words or pictures; called targets) that are affectively congruent with the priming stimulus. For example, we are faster to judge the valence of the target word *love* when this is preceded by an affectively congruent prime word (e. g., *dolphin*) rather than an affectively incongruent one (e. g., *ugly*). This effect seems even present for subliminally presented primes (Hermans, Spruyt, De Houwer, & Eelen, 2003), but is short-lived, with most robust effects found for short stimulus onset asynchronies well below 300 ms (Klauer & Musch, 2003).

We reasoned that if indeed solving a two-tone image is intrinsically positively valenced then we might be able to pick up their priming effect on response times for subsequent word evaluation. Faster responses to positive (relative to negative) word targets after presentation with solved two-tone images would imply a more positive valence of these stimuli. Importantly, solving the two-tone images was no explicit part of the task, only the rapid affective categorization was. This reduced the likelihood that any affect we might find is attributable to performance affect of doing the task well (such as pride).

4.1 METHOD

4.1.1 *Participants*

44 students (14 men) at the University of Leuven participated in the experiment, after we obtained their informed consent. They all had normal or corrected-to-normal vision and were native speakers of Dutch. They were either paid or received course credits for their participation.

Positive target words:	plezier, zomer, liefde, verjaardag, eerlijk, blij, moeder, vlinder, vrede, zon, boeket, humor, wens, bloesem, kus, lente, vogel, eendje, katje, hert, panda, dolfijn, vriend, zwaan, peuter, broer, genie, moeder, optimist, meisje, held, expert, poesje, baby, bruid, dolfijn, kerstboom, teddybeer, bloem, pup, zwaluw
Negative target words:	lelijk, kanker, geweld, moord, plaag, afval, vijandig, ziekte, vuilnis, angst, paniek, dood, ongeluk, puist, vet, verdriet, pijn, ratten, gier, slang, muis, mug, kakkerlak, worm, tiran, gangster, bruut, snob, egoist, beul, dief, sadist, hoer, vijand, wormen, lijk, spin, explosie, schedels, pistool, hitler

Table 1: All positive and negative target words used in the experiment. Selected from Hermans and De Houwer (1994)

4.1.2 *Stimuli*

The two-tone stimuli are created by smoothing (low-pass filtering) and thresholding grayscale images such that brightness values below a certain threshold are plain white, and those above are black. We used a set of two-tone images with accompanying gray-scale solutions from Imamoglu, Kahnt, Koch, and Haynes (2012). Based on a pretest with 20 participants (different from those in the actual experiment), we selected 85 (from a set of 122 two-tone images in total) by removing those that were immediately recognized and those that, even after providing the grayscale solution, were not clear. We generally wanted to avoid that participants could solve the two-tone images on their own. Note, however that this cannot be objectively

determined for a given image, because of large individual differences (Ludmer et al., 2011). The prime pictures all portrayed single identifiable objects in the foreground, but we took care not to include ones that comprised clearly emotional objects.

The positive and negative word stimuli that served as targets were taken from a normative study using affective ratings for Dutch words (Hermans & De Houwer, 1994) and have been successfully used in several affective priming studies (e. g., De Houwer, Hermans, Rothermund, & Wentura, 2002; Spruyt, De Houwer, Hermans, & Eelen, 2007). Table 1 provides a complete list of all words used (41 positive and 41 negative). In each trial, a word was randomly chosen (without replacement) from the list of possible words, depending on the condition. A *Python* program (using the *Psychopy* library; Peirce, 2007) controlled stimulus presentation and response collection.

4.1.3 *Procedure*

Participants were seated in a dimly lit room, with their heads placed on a chin rest to keep a constant distance to the screen. The experiment consisted of two phases, though crucially, participants were not informed about the second phase beforehand. In the first phase (affective priming test) participants had to categorize positive or negative words after being briefly presented with the solution of a two-tone image. All participants first received ten practice trials (with images that were different from those in the actual experiment) to familiarize them with the trial structure. The procedure of phase 1 is depicted in Figure 17. Participants were first presented with a two-tone image for 3.5 s, subsequently they briefly saw the corresponding solution (500 ms), and then the two-tone again (200 ms). Returning to the two-tone image after the solution served to enhance the aha-experience because it made clear how the back-and-white surfaces actually fitted together. After 200 ms, a positive or negative word appeared in red on a white rectangle, superpositioned on the center of the two-tone image. Participants were instructed to respond as quickly as possible with the left or right mouse button. We counterbalanced the mapping of left and right buttons to positive and negative evaluation between participants. An empty screen remained on until the participant responded, after which a new trial sequence started (randomized order).

Based on introspection, we assume the predictive progress and hence the positive affect when viewing two-tone images is largest for the first presentation. Indeed, the big insight moment can hardly

be repeated ('one-shot' learning). Hence, we did not repeat the images. This meant we could use only 85 trials per participant.

In the second, unannounced phase (recognition test), participants were asked to try and solve the two-tone images that were used in the first phase. Here, each two-tone stimulus was again shown (for 2 s) in a random order and participants were instructed to type in one word that describes what they thought was the actual content of the two-tone image. They were encouraged to guess, but could just continue if they were unable to recognize anything.

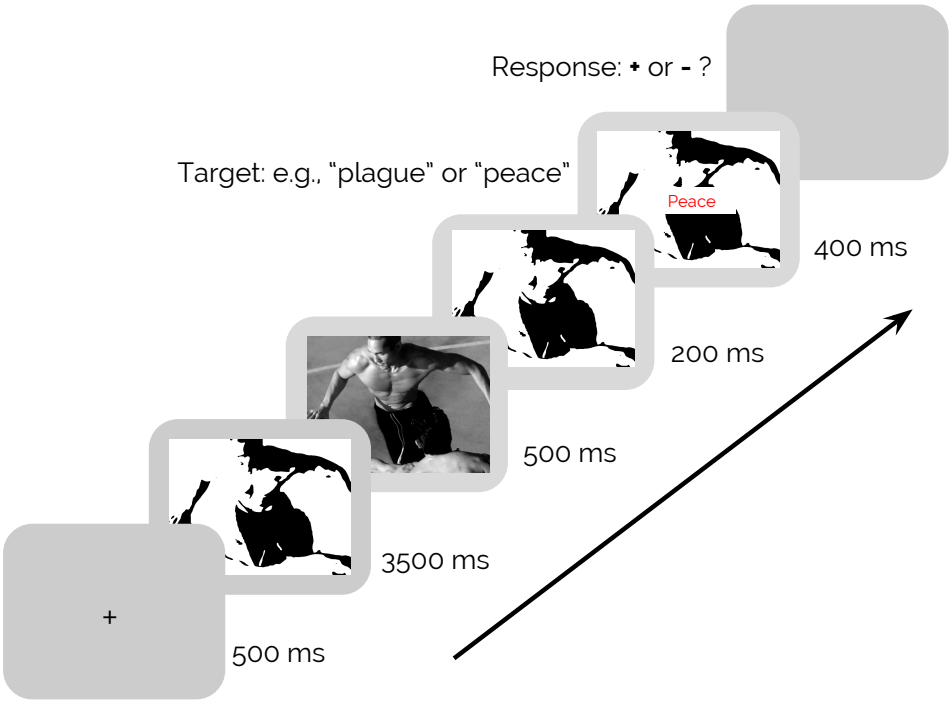

Figure 17: Procedure of the affective priming phase.

4.2 RESULTS

We used the recognition performance of phase two (a binary variable, henceforth 'recognition') as a measure of whether participants actually solved the two tone image during the first phase (i. e., saw the link between the two-tone image and the grayscale solution). If they were able to do so, we predicted they would respond faster to positive than to negative target words. Hence, our main interest is the interaction effect of valence and recognition on reaction times for the valence categorization task. But before we get to that, we discuss a few complementary statistics of the data. Four subjects had to be removed from all analyses because of errors in the data collection (1), extremely long reaction times (2) or non-compliance

to instructions (1). All analyses were executed in *R* (R Core Team, 2014) with the *lme4* package (Bates, Mächler, & Bolker, 2012).

We first fitted a mixed logistic regression model on categorization performance (of word valence) with participant as random effect, and word valence and recognition (in phase 2) as fixed effects. Both the main effect of recognition (z=-2.1, p= 0.04) and the interaction of valence with recognition (z=2.56, p= 0.01) were significant, while the main effect of valence was not significant (z= 0.79, p= 0.43). Looking at Figure 18, we can see that only for negative words categorization accuracy seems somewhat better if the preceding image later was not recognized, than when it was recognized. This may indicate that evaluating negative words, while not significantly more difficult than positive words, is more affected by processing of the previous images: the more attention stays on those images, the worse categorization of negative words. Note that this happens without an instructed task on the two-tone images and that the solution was never present simultaneously with the word to be categorized. More generally, an error rate of 5-10% on categorization responses possibly indicates that the some of words were not unambiguously positive or negative for all participants.

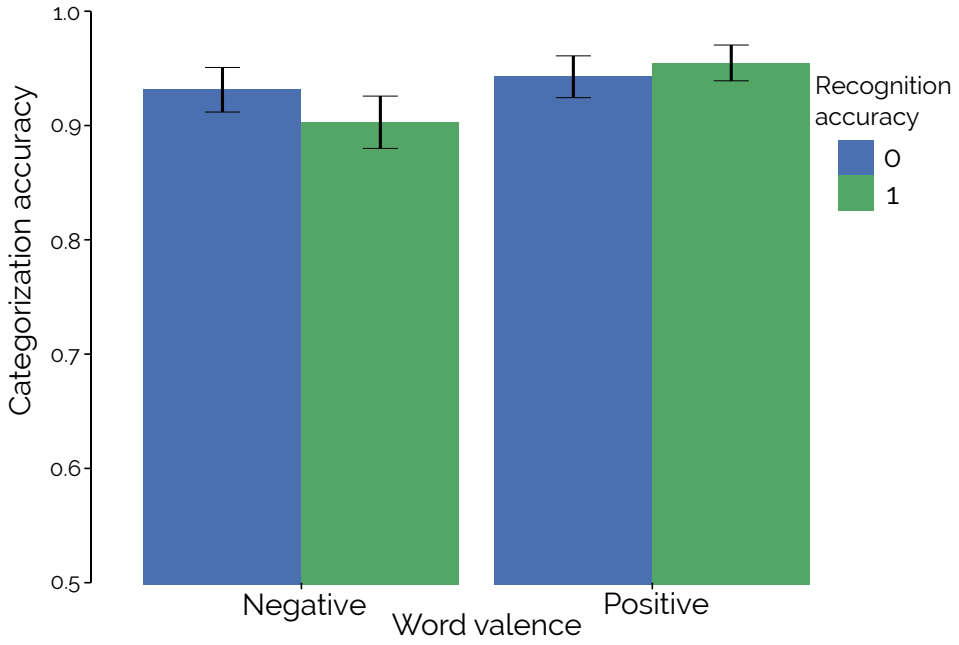

Figure 18: Accuracy of categorization responses for positive and negative target words, by recognition. Error bars are corrected, within-subjects 95 % confidence intervals (not standard error) as described by (Morey, 2008).

For the remaining analyses we only included correct word categorizations (93 %). Reaction times of the affective priming test that

were below 200 or above 2000 ms were considered lapses or invalid responses and hence removed from analyses. Subsequently, reaction times further than three times the standard deviation away from the mean of this individual were removed as outliers (in total 1.6%). We also verified whether we have approximately the same accuracy of recognition in phase 2 (so equal number of data points) for positive and negative words (phase 1). Mean recognition rate in the second phase was 49%, and did not significantly differ depending whether it was followed by positive or negative target word in the first phase (t(3082.35)=-0.78, p= 0.44). Hence we have a practically equal amount of recognized and unrecognized trials, for both the positive and negative words.

Next, we quantified the degree of consistency in recognition for subjects regarding the different two-tone images with Fleiss' Kappa (Fleiss, 1971; Gamer, Lemon, Fellows, & Singh, 2012), which was low (κ = 0.13), signifying only slight 'agreement'. This means that the set of images that were correctly recognized (i. e., elicited 'insight'), differed between subjects, consistent with earlier reports (Ludmer et al., 2011) and with our intuition that not all solved two-tone images are equally compelling to everyone. The 'goodness' of a solution very much varies across images and across individuals.

Finally, we fitted a mixed linear model on the reaction times (phase 1) with word valence and recognition as fixed effects, and participant as random effect. Figure 19 depicts these comparisons. At first sight, people seem to be generally quicker to respond to positive than to negative words, but this seems primarily the case when primed with subsequently recognized images, consistent with our hypothesis. However, the full fitted model suggests otherwise. First, the random participant effect is highly significant ($\chi^2(1)$= 1320, p<0.0001) as one might expect. Second, neither valence (t(3043.3)= -1.66, p= 0.097), nor recognition (t(3049.2)= -0.27, p= 0.78), nor the interaction of valence and recognition (t(3043.9) = -1.57, p= 0.12) reaches significance. We decided to split the dataset in two, fitting a mixed linear model for primes that were recognized and one for unrecognized primes separately. For primes that were later recognized there was a significantly faster response to positive words than to negative words, consistent with our hypothesis (estimated difference=-32.22, t(1481.7)= -3.96, p<0.001; $M_{negative}$=801.7, 95% confidence interval= [754,850]; $M_{positive}$=769.4, 95% confidence interval= [722,817]). However, for primes that were not recognized the effect of word valence on reaction times did not reach significance (estimated difference=-12.137, t(1526.7)= -1.33, p=0.18). This way of subdividing data and comparing significance levels of separate tests is a very common practice, but, as Nieuwenhuis, Forstmann, and Wa-

genmakers (2011) point out, it does not actually give a valid answer if one is interested in the interaction, as we are here. Simply put, "the difference between significant and not significant need not itself be significant" (Nieuwenhuis et al., 2011, p. 1105). Hence, we only have suggestive, no conclusive evidence for the difference we were getting at.

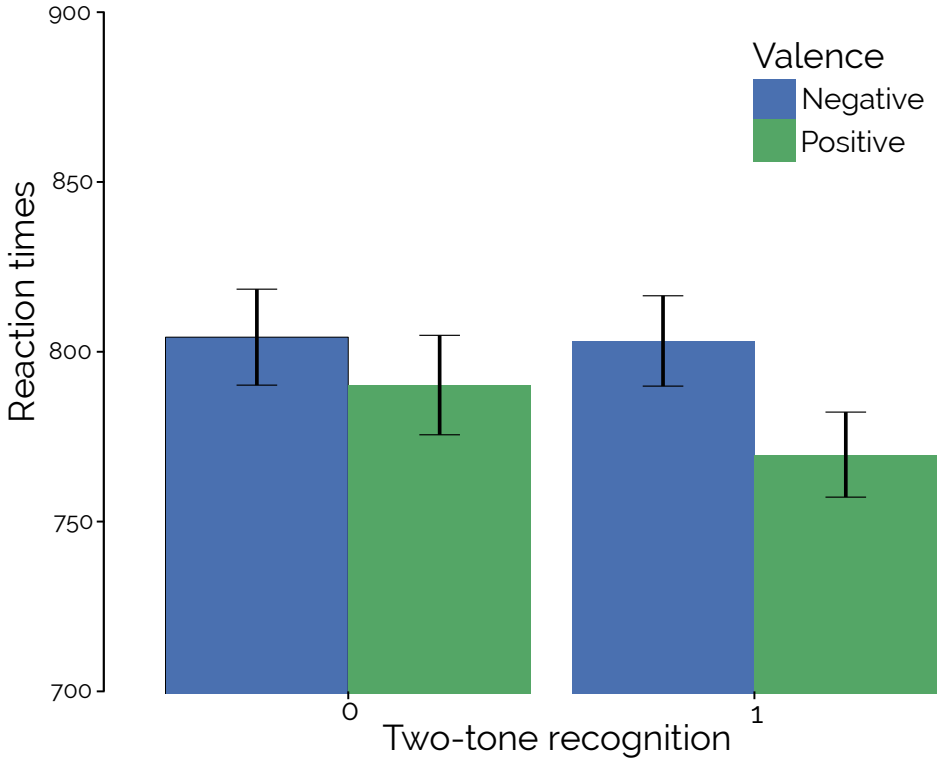

Figure 19: Reaction times for positive and negative target words, by recognition accuracy. Error bars are corrected, within-subjects 95 % confidence intervals (not standard error) as described by Morey (2008).

4.3 DISCUSSION

Our findings provide some evidence for the hypothesis that reorganization of perceptual input is marked by positive affect. We suggest that this positive experience, often called perceptual insight, is specifically linked to the progression from unpredictable to familiar and organized percepts, which we incorporated in our study by using two-tone images and their solutions.

One neuroimaging study also provides some indirect but suggestive evidence for the affective nature of solving a two-tone image. Ludmer et al. (2011) found that amygdala activity during the 'in-

sight moment' reliably predicted memory for the specific content of the image (given only the degraded image) a week later. The insight moment was elicited by presenting people briefly with the full gray-scale solution of the two-tone. The higher amygdala activity during one of these moments, the more likely the corresponding solution was correctly remembered a week later. No behavioral measure of affect was used in this study, so amygdala activity could in principle represent negative or positive emotion or even a neutral kind of 'relevance' signal (Sander et al., 2003). Importantly, the amygdala is known to be also involved in reward and positive emotions (E. A. Murray, 2007). Our findings indeed suggest that specifically positive affect is elicited in these instances.

With regard to the relation between uncertainty (prediction error) and positive affect, three studies that are conceptually related to ours, deserve mentioning. In the first (Jepma, Verdonschot, van Steenbergen, Rombouts, & Nieuwenhuis, 2012), participants passively view drawings of everyday objects that were preceded by a heavily blurred, irrecognizable version of either the same object or another, unrelated one. The idea is that if the first image portrayed the same but blurred object, there is a clear reduction in uncertainty when the second (well recognizable) image come on, so a more positive appreciation. Indeed, higher fMRI activity was found in striatal areas usually associated with reward processing for this condition (compared to the nonmatching condition). Winkielman and Cacioppo (2001), on the other hand, briefly primed everyday objects with their own outline (internal details removed) or those from a different object and found more zygomaticus activity ('smiling muscle') for the congruent primes. This is similarly consistent with the uncertainty reduction hypothesis.

A last relevant study (Schouppe et al., 2014) first replicates findings from Dreisbach and Fischer (2012). The latter demonstrated facilitation of categorization responses to negative (vs. positive) stimuli after presentation of incongruent (vs congruent) Stroop stimuli, implying that cognitive conflict or prediction error in these stimuli is negatively valenced. Importantly, the Stroop stimuli were not part of the task (note the parallel with our study), and so did not need to be solved. Schouppe et al. (2014) corroborated these findings in their first experiment, but in the second they did ask participants to respond to the Stroop stimuli. What they found was that incongruent, solved primes induced faster responding to positive (relative to negative) targets, presumably because the conflict was resolved, i. e., initial prediction errors reduced. Clearly, this is consistent with our hypothesis, although positive affect associated with good performance for difficult trials (e. g., pride) cannot be excluded here.

Still, the fast time scale of these effect suggests an automatic rather than a reflectively mediated mechanism, possibly indicating that this affect is more fundamental than a complex, social emotion like pride. In our study, we tried to de-emphasize solving two-tone as a task, to allow for as pure as possible a measure of intrinsic affect of the perceptual transition, but it is entirely possible that participants made this into a task. Future studies may experiment with using a main task with a higher cognitive load than our reaction time task to avoid this, but this may have the danger that background two-tone and solution are not processed at all. Another variation would be to explicitly ask subjects to solve the two-tones (i. e., a true dual task setting), and inform them in advance about the recognition phase, to see if positive affect is indeed intensified (compared to the no two-tone task condition) for correctly recognized images, as we presumed.

We should be attentive to other potential limitations and alternative interpretations of our findings. Particularly, our findings may be compatible with the perceptual fluency account of (aesthetic) appreciation. Perceptual fluency denotes the ease with which stimulus material is processed (Reber et al., 2004; Reber, Winkielman, & Schwarz, 1998). Although it could favorably be interpreted to be about processing *dynamics* analogous to our idea of predictive progress (see Chapter 6), its operationalizations so far have been rather static, namely in terms of prototypicality, familiarity, symmetry, clarity, etc. The success of the fluency concept is also its biggest limitation: it is atheoretical in the sense it is not grounded in a plausible theory of general cognitive processing. It is about processing fluency without looking at how this processing actually happens. Hence, it relies on our intuition of what is 'fluent', and this vagueness opened the way to a wealth of studies manipulating different alleged forms of fluency. Most, if not all, of them can be reinterpreted in terms of predictability, but rarely dynamics are explicitly considered. We have tried to build the dynamics into our experimental paradigm, but our results are still consistent with the fluency account if fluency is taken to mean the experience of rapidly reducing uncertainty (prediction errors). However, in our view the high unpredictability of the starting point (two-tone) is instrumental for the later positive affect. We plan to empirically draw this distinction more clearly either by using non-matching (inverted) two-tone images before the solutions or by using "multiple-tone" images. The former has the drawback that it may induce negative affect related to the inconsistency (prediction error). The multiple-tone images may be better controls, because they are distorted but still immediately recognizable without solution (no transition).

One could construct one other alternative explanation for the current findings, by inverting the causality we assumed in our explanation. It is possible that positive words are easier to process and hence detract less resources from the processing of the solutions of the two-tone images that came just before (though it disappeared before the word came on). The easier the word is processed (according to the reaction times), the better the image is later remembered. Equal recognition scores for images that previously preceded positively valenced words versus negatively valenced words speak against this explanation, but strictly we cannot exclude it.

Finally, our pilots suggest that timing is crucial, indicating that these effects are short-lived. This is known for conventional affective priming, where time course (stimulus onset asynchrony) was already systematically investigated (Hermans, De Houwer, & Eelen, 2001). Our study, however, provides evidence for a form of 'affective' priming using (semantically) neutral primes, which probably means the effects are even subtler. An additional complication in our setting is whether positive affect is strongest from the moment the solution is presented or only after, when the two-tone is back on and its pieces fit together with the newly gained knowledge. Based on introspection we opted for the latter in the current experiment, but this should be more systematically examined.

These timing issues, together with the individual differences in (image-dependent) compellingness, and the fact that we could not repeat trials (images) leads to very noisy measurements for already subtle effects (less than 20 ms), which likely explains the non-significance of the interaction in the full model.

Although adequate, quantitative measures of economy of representations are still lacking, we could plausibly attribute the positive affect we found to the internal progress made in representing the input in a sparse and efficient way (see also Ludmer et al., 2011). The newly found structure can summarize the disparate inputs best and accurately predicts the spatial configuration of black and white surfaces in the two-tone image. Indeed, it can be shown that minimizing prediction errors as described by the predictive coding model, is equivalent to finding the so-called minimum description length (given Gaussian noise; Huang & Rao, 2011). The more predictable patterns can be found in input data, the more compressible it is (Schmidhuber, 2010). In this sense, predictive coding converges on sparse, simple representations given the inputs to be explained. Settling on a model —learned by prior experience, in this case: the solution of the two-tone image— that allows sudden predictive progress seems marked by positive affect. Moreover, the suddenness or unex-

pectedness of this reduction of prediction errors seems to intensify the pleasure (for a possible explanation, see Chapter 6).

That aesthetic pleasure has much to do with frugal use of energetic resources, was also clear to Eysenck when he formulated his prescient 'law of aesthetic appreciation': "The pleasure derived from a percept as such is directly proportional to the decrease of energy capable of doing work in the total nervous system, as compared with the original state of the whole system." (Eysenck, 1942, p. 358). In turn, Eysenck acknowledged his indebtedness to Köhler and Koffka. The former writes: "in all processes which terminate in time-independent states, the distribution shifts towards a minimum of energy" (Köhler, 1920, p. 250).

Three things stand out in Eysenck's law. First, it emphasizes that pleasure is proportional to a *change* ('decrease'), consistent with our hypothesis. Second, it remains vague on how to relate 'energy' to cognitive variables. One page earlier, Eysenck writes: "those external stimuli will be judged the most beautiful which are most in agreement with the internal forces of perception." (Eysenck, 1942, p. 357). Again, what those internal forces are is not well specified. We can make an attempt to fill this in using predictive coding. These internal forces would then be equated with the models (priors or predictions) one's perceptual system has about the world. If the inputs match these predictions, positive appreciation results. Importantly, when predictions (models) can be varied and hierarchical, 'objectively' complex stimuli can become pleasing. However, so far this is a static view, omitting the decrease that is crucial according to the law. Clearly, to align our hypothesis with Eysenck's, the 'energy capable of doing work' should be proportional to prediction error, such that a decrease in energy capable of doing work is equivalent to a decrease in prediction errors. The energy capable of doing work is also called Helmholtz free energy. Note that prediction error minimization can be generalized into the free energy minimization principle (free energy is an upper bound on surprisal or negative log probability of sensory data given a model; Friston, 2010b). However, this free energy is an information-theoretic quantity first used in statistical physics, only called free energy because of its formal similarity (common probabilistic basis) with Helmholtz free energy. Still, as Sengupta, Stemmler, and Friston (2013) prove, there is a close link between the two. The full argument is rather technical, but the basic conclusion is that "commonly occurring representational states —that are a priori most probable— are the least [metabolically] costly" (Sengupta et al., 2013, p. 9). This implies that unlikely or unpredicted (surprising) inputs, i. e., with lots of prediction errors, have high metabolic cost. Helmholtz free energy and information-

theoretic free energy share the same minimum (equilibrium). An interesting re-evaluation awaits here of old Gestalt ideas (e. g., the minimum energy idea, isomorphism, etc.) in the light of these new developments linking thermodynamics to informational concepts of order (predictability) and complexity. Although the details differ from what Gestaltists proposed, the basic intuition about the connection between Gestalt formation and thermodynamic, bioenergetic processes may still be validated. Such understanding would also make the equivalence more explicit of Eysenck's law and our hypothesis on positive affect and prediction error decrease.

Just reasoning from predictive coding also provides an intuitive answer with regard to reduced processing demands. If, as the predictive coding scheme prescribes, a) the brain explicitly represents prediction errors as activity of separable neural units, b) it works by minimizing prediction errors, then neural activity (and hence metabolic work) decreases as the brain gets better at predicting its environment (reducing prediction errors; FitzGerald, Dolan, & Friston, 2014). Increased predictive power directly implies decreased metabolic use. However, for the net result to be a decrease in cost, improved predictions, that also need to be represented, have to be more efficient than the input activity they can suppress. This is a reasonable assumption because the generative model (predictions) can be coded by changes in synaptic efficiency (Friston, 2003). Studies have shown that grouping or the formation of a Gestalt is linked to reductions in fMRI activity in the early visual cortex (Fang, Kersten, & Murray, 2008; S. O. Murray et al., 2002), suggesting that indeed top-down predictions can explain away matching inputs (but see de-Wit, Kubilius, Wagemans, & Beeck, 2012), hence reducing processing demands. However, both of these studies also report concurrent activity *increases* higher up in the visual hierarchy, in object-sensitive regions. This seems to go against a more efficient representation, because it suggests processing resources are just redistributed upwards (not reduced). Still, predictive coding hypothesizes that across the full hierarchy of models in the brain, including the two purported levels considered by these studies, prediction errors (and hence metabolic cost) are minimized during perception.

This brings us seamlessly to the third important thing in Eysenck's law: it talks about a decrease relative to the original state of the 'total nervous system'. Considering prediction errors (metabolic demands) in the whole system, seems indeed important in the context of (art) appreciation. Situation and context, to the extent they shape pre-stimulus brain activity, could determine (dis)likings of art or other perceptual stimuli, although this is a challenging proposition to investigate systematically. It might explain why people often

like simple, stereotypical stimuli in certain settings (e. g., unfamiliar ones), but more complex, challenging stimuli in other settings. In any case, Eysenck's law combined with predictive coding seems a promising starting point to further examine the link between liking and thriftiness in the brain.

In conclusion, manipulating processing dynamics in terms of prediction errors present in the (perceptual) system is challenging at this stage, because we lack the means to directly probe them. Neural measures would not provide much solace, until we can verify that, as predictive coding holds, dissociable neural units encode predictions and prediction errors at different levels of the perceptual hierarchy (de Gardelle et al., 2012). In the meantime, our study suggests that dynamics in predictability can be used in the lab and may have intrinsic affective consequences. Specifically, we have gathered some support for the idea that insight in two-tone images as measured by later recognition is linked to positive affect, i. e., faster categorization responses on positive relative to negative words. We interpret this in terms of the predictive progress viewers make going from two-tone images to their solutions. However, given that the statistics were not unequivocal, this effect needs to be replicated first, preferably including the indicated controls.

We do not see things as they are but as we are.

— Anaïs Nin, 1961

5

Can affective relevance bias bistable perception?

For scientists, the question of whether the quote above could literally be true is a fascinating, albeit a thorny one. Surely perception unchecked by external reality (i.e. hallucination) would serve us poorly. But when allowed only a glimpse, a perception biased by our own fears or hopes might enable faster or more appropriate responses. Theoretically, a role of emotion in perception has become more plausible since it has become clear that seeing relies heavily on top-down information flows, which include affective context (Barrett & Bar, 2009). Empirically, two opposing patterns of results have emerged. On the one hand, several studies find a negativity bias: a bias towards negative or threatening emotional stimuli, including the large body of work on attentional bias for negative facial expressions (Bar-Haim, Lamy, Pergamin, Bakermans-Kranenburg, & van IJzendoorn, 2007) and on better detection of negative stimuli (e. g., Dijksterhuis & Aarts, 2003). Reports by Stefanucci, Proffitt, Clore, and Parekh (2008) that acrophobic volunteers estimated vertical distances to be greater also fit within this negativity bias. Finally, negative emotional faces seem to break through more easily in continuous flash suppression (Yang, Zald, & Blake, 2007) and in binocular rivalry (Alpers & Gerdes, 2007), although for the latter a general emotional enhancement, irrespective of valence, is also reported.

By contrast, a second line of studies reports a positivity bias, also called wishful seeing. This line dates back to the 1940s, when it was found that observers tended to perceive the interpretation of an ambiguous figure that was previously rewarded (e. g., Schafer & Murphy, 1943). This tradition has received renewed attention recently with two studies confirming a role for motivation in perception. One used distance estimation and a throwing task to find that desired objects are seen to be closer (Balcetis & Dunning, 2010), while another showed that the first percept we experience for a bistable figure (e. g., 13 vs. B) can be influenced by what we currently prefer to see (Dunning & Balcetis, 2013). Finally, Voss, Rothermund, and Brandtstädter (2008) rewarded or punished ambiguous color patches differently and found that positive or non-negative stimuli required

less information to be classified and were processed faster. Hence, the enhanced perception of positive perceptual alternatives seems to be more than just a post-perceptual decision bias.

Apparently, it can be hard to predict what type of bias (positive or negative), if any, will emerge in a particular perceptual situation involving emotional stimuli. Indeed, task context and even personality traits may play a crucial role. For example, trait emotion could be an important modulating factor, but most studies only looked at stimulus-emotion or short-term induced emotion. Gray, Adams, and Garner (2009) did look at trait anxiety in relation to binocular rivalry with faces and found that highly anxious individuals tend to perceive angry and fearful faces as more dominant, consistent with a negativity bias. Additionally, methodological concerns hamper the evaluation of the available evidence. The studies using distance estimation are prone to post-perceptual decision biases, while in the binocular rivalry studies there is the additional problem of low-level differences between neutral and emotional stimuli (e. g., faces) which could cause the observed differences in perception. We used a different bistable stimulus, an ambiguous point-light walker, to accommodate for these shortcomings.

Vision scientists have embraced biological motion stimuli, because of the balance they strike between fine manipulability and immediate social and ecological relevance. Although these figures are constructed from only a handful of moving dots placed on the joints of a moving actor, viewers easily recognize their gender, emotions and intentions based on gait dynamics (for a review, see Blake & Shiffrar, 2007). Fewer studies explored their potential bistability, first noticed by Vanrie, Dekeyser, and Verfaillie (2004). When projected without perspective information (orthographic projection), a walking figure facing the viewer can just as well be seen as facing away from the viewer, a categorically different percept (Figure 20). Both interpretations are anatomically plausible and in principle equally likely. Importantly, low-level input characteristics remain exactly the same for both percepts. It turns out, however, that people perceive the walker in about 80% of the cases as facing them (Vanrie et al., 2004). The social or biological relevance of a person facing you is considered to be at least partly responsible for this so-called facing bias (Brooks et al., 2008; Vanrie et al., 2004). Indeed, the cost of not detecting an approaching person is potentially much higher than that of a false alarm.

Although social anxiety disorder (social phobia) only appeared in the Diagnostic and Statistical Manual of Mental Disorders (DSM-III) in 1980, it is today one of the most common psychiatric disorders. The 12 month prevalence in the general population is 15.6%, but

signs of social anxiety as a personality trait (shyness, fear of public speaking, social avoidance) are widespread in healthy populations (Furmark et al., 1999). Assuming that for people with high social anxiety the difference in relevance of the two percepts of the biological walker is even more pronounced, we conjectured that this would be reflected in how these people perceive the walkers. In the real world, a person walking up to you implies an imminent social interaction with this agent. Note, however, that the body is reduced to a few dots in these stimuli, and the face is completely absent (one dot). Hence, these walkers are objectively neutral and any bias will be in the eye of the beholder. Since no fear-inducing cues (e. g.a facing face) are present, they can be considered to be the cleanest test for a fear of approach as such.

Based on the studies finding a stronger attentional and interpretational bias in phobics towards objects of their fears (Bar-Haim et al., 2007), and one report of increased dominance of negative faces in binocular rivalry for anxious people (Gray et al., 2009), we deemed it likely that a similar tendency would hold for our bistable stimuli. Namely, that people with high social anxiety would perceive the walker in its more threatening, approaching configuration and thus would report more facing towards viewer percepts, compared to non-anxious subjects. Still, we acknowledged from the outset that our stimuli were in several ways quite different from those previously used, most importantly that there was no explicit emotional manipulation within them. We were also aware of the studies finding a positivity bias, which led to the formulation an alternative hypothesis that predicts what one could call a self-serving bias in perception: a tendency to perceive the safer configuration of a person walking away from you. The latter bias could come about by an active enhancement of the more positive percept signaling no social interaction, or by an active avoidance of the more negative percept signaling a future social interaction, consistent with what is found for social stimuli in daily life in social anxiety.

5.1 METHOD

5.1.1 *Participants*

On the basis of a validated, reliable questionnaire for social anxiety (Liebowitz Social Anxiety Scale; Fresco et al., 2001; Liebowitz, 1987) we selected high and low socially anxious participants from a pool of 450 first bachelor psychology students. People with low anxiety had a total score lower than 25 (percentile .25, n = 19, M = 15.3, SD = 4.53, all females), while those with high social anxiety had a score higher than 65 (percentile .88, n = 18, M = 76.8, SD = 8.02, all

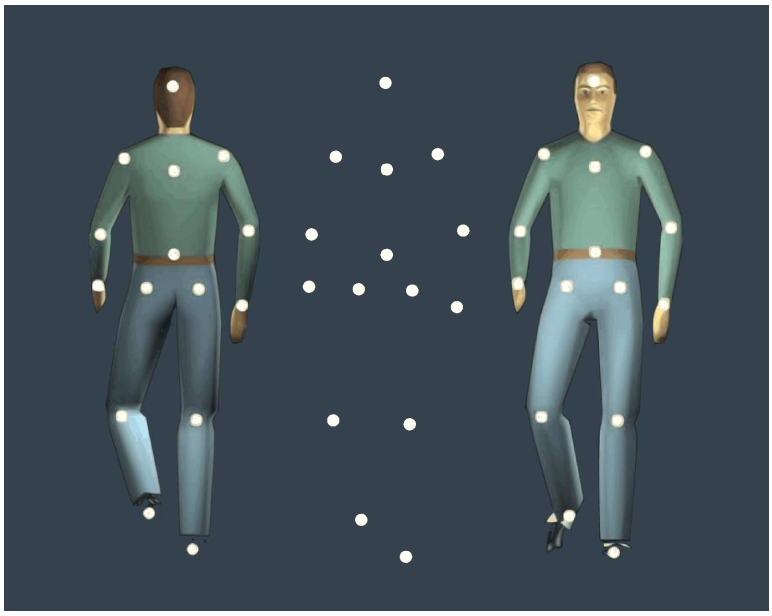

Figure 20: Point-light stimulus (middle) flanked by overlays illustrating the two possible interpretations.

females). The cutoffs were chosen to closely match the normative study of Fresco et al. (2001), in which non-anxious controls had a mean score of 14.5, while the patients diagnosed with social anxiety disorder had a mean score of 74.5 on the questionnaire. To avoid artifacts and to exclude clinical anxiety, people with very low (<5) and very high (>85) scores were excluded. A brief questionnaire after the actual experiment confirmed that none of the participants were diagnosed with a clinical mental disorder. They received course credit for their participation. All the participants had normal or corrected-to-normal vision. The study was approved by the Ethical Committee of the Faculty of Psychology and Educational Sciences of the University of Leuven. Written informed consent was obtained from all participants.

5.1.2 *Stimuli and procedure*

Following a procedure developed by Schouten and Verfaillie (2010), we parametrically varied the amount of perspective information in the point-light walkers to systematically measure the bias for each participant. In practice, this boils down to disambiguating the walker by adding small amounts of perspective information of a walker that either approaches or walks away. In perspective projection a change in the distance between the collection of dots and the projection plane (or equivalently the field of view angle) causes a

change in the relative locations of the 2-D projections of the dots on the display (for more details, see Schouten & Verfaillie, 2010). A period of two to four months separated the screening from the actual experiment, which made the link less obvious. Participants were not informed about their social anxiety score, and the experimenter was blind to the social anxiety group the participants belonged to. In a dimly lit, soundproof room the participants were randomly presented with a point-light walker (15 dots placed on coordinates from Troje, 2002; 8°of visual angle) of one out of 13 different levels of perspective information, for a total of 520 trials (40 repetitions per level). Observers had to respond with the up and down arrows of the keyboard to indicate whether they saw the walker as facing towards or away from them (2-alternative forced choice). The walker remained (moving) on screen till the subject responded (no time limit). We instructed participants to focus on the center of the stimulus throughout the presentation and to respond according to their first impression.

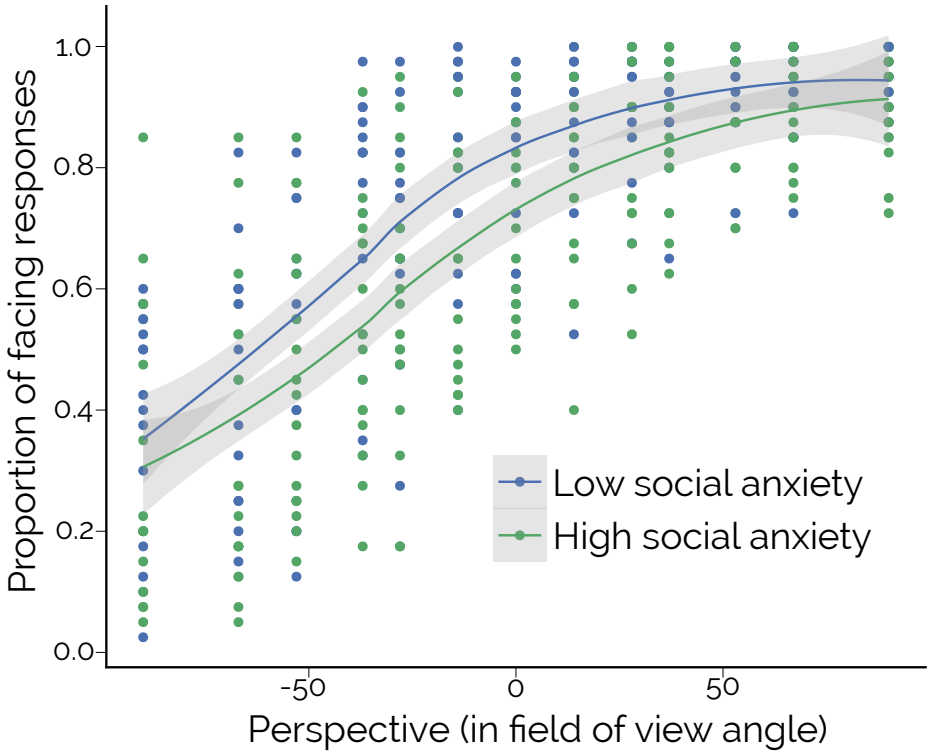

Figure 21: Mean proportion of facing towards viewer responses by amount of perspective information for the high social anxiety group (green) and low social anxiety group (blue). Curves represent local regression fits (LOESS). Shaded area is standard error of mean.

5.2 RESULTS

As already mentioned, people generally have a strong facing bias, with some individuals showing no variation at all in their experience even when presented with walkers with different amounts of perspective information. Participants that reported seeing the walker in more than 90% of the cases as coming towards them for all levels of perspective information were excluded from the dataset (5 in total, 3 of the high social anxiety group). However, analysis of the full set did not influence the main conclusions. We applied a two-way ANOVA on the probit transformed proportions of facing towards viewer responses with anxiety group (between-subjects) and perspective information (within-subjects) as factors. As expected, the amount of perspective information greatly influenced perceived orientation ($F(1,414) = 451.4$, $p<.001$), confirming the effectiveness of the parametrical perspective manipulation. The main effect of group was also significant ($F(1,414) = 11.4$, $p<.001$), with high social anxiety observers systematically showing lower proportions of facing towards viewer responses compared to low social anxiety observers (Figure 21).

Though it was not of primary interest, we also performed a repeated measures ANOVA on the reaction times (Figure 22). Here the influence of perspective information was significant too ($F(1,30) = 21.8$, $p<.001$), indicating that less perspective information creates longer reaction times. This may be related to increased processing necessary to settle on a stable percept for more ambiguous figures. No main effect of group ($F(1,30) = 0.6$, $p = 0.45$) was present, but there was a significant interaction of group and perspective information ($F(1,30) = 7.8$, $p<0.005$), apparently due to higher reaction times of the low social anxiety group on the first three levels of perspective information. This overlaps but does not coincide with the clearest discrepancies between groups on the facing responses.

5.3 DISCUSSION

Our findings confirm that trait anxiety has an effect on the perceived in-depth orientation of point-light figures. Contrary to our original hypothesis, we found that higher social anxiety is associated with higher proportions of facing away percepts. We did not observe a greater dominance of the more negative (threatening) interpretation, but rather a suppression of the negatively valenced percept, an enhancement of the more positive configuration, or a combination of both. However, as discussed in the introduction, such a positivity bias has been reported earlier in studies of motivational effects on

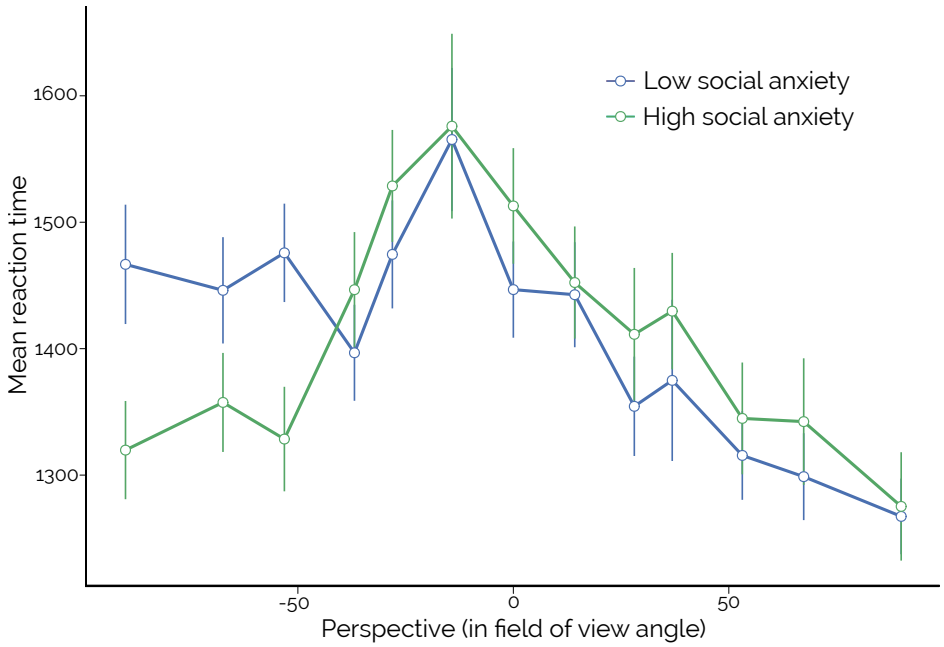

Figure 22: Mean reaction times for the high social anxiety group (green) and low social anxiety group (blue). Error bars represent standard error of mean.

perception (Dunning & Balcetis, 2013; Voss et al., 2008), but not yet for socially anxious participants. Similarly to Voss et al. (2008), our highly anxious participants may have processed the non-negative 'facing away' interpretation more quickly, giving it the necessary edge over the competing facing-towards representation in the race for awareness. Alternatively, an active avoidance of the more negative percept signaling social interaction could lead to a similar outcome. Avoidance of threatening stimuli is frequently reported in the behavioral literature on social anxiety (Bögels & Mansell, 2004). Even attention studies often report an avoidance of emotional stimuli immediately following the increased capture of attention, prompting researchers to propose the vigilance-avoidance model of attention in social anxiety (Bar-Haim et al., 2007). In daily life, avoidance of social interaction is a major impairing factor for people with social anxiety and in the screening questionnaire we used, half of the questions concerns avoidance of social situations. In our experiment there was always a possible interpretation avoiding the approaching figure, and this may have been implicitly prioritized by people with social anxiety.

Both explanations imply that top-down information in the form of trait can bias perceptual organization. This is consistent with earlier findings that the meaning of a perceptual organization can in-

fluence the formation of this organization in a top-down manner (e. g., Peterson, 1994). Additionally, our study suggests that the facing-towards-viewer-bias in the general (non-anxious) population (Vanrie et al., 2004) is not due to anxiety elicited by the facing interpretation of these stimuli. Indeed, if this would be the case, selecting people with higher anxiety would produce an even stronger facing bias (the inverse of our findings). Assuming that the biological relevance of the facing percept causes the facing bias, this effect is probably not mediated by anxiety (i. e.the threat value of an approaching percept).

Nonetheless, some caution is needed when interpreting the current data, since most previous studies in clinically anxious groups found a negativity bias, both in attention for threat-relevant stimuli and in the interpretation of ambiguous stimuli. Why did we not observe such a negativity bias? First, it is important to realize that our task is not strictly speaking an attentional one. There was no right or wrong in our task, so little or no performance context, which may influence people with high and low anxiety differently. Also, most reports of an attentional bias towards negative, fear-relevant stimuli used clinically anxious groups (Bar-Haim et al., 2007) while we only had nonclinical highly anxious participants. The perceptual and attentional strategies of clinical groups could be special in their focus on negative stimuli. Additionally, most research reporting a bias towards social threat in high social anxiety uses faces, the competing alternative being another, neutral face. According to some studies, these neutral faces are experienced as negative in high social anxiety (e. g., Cooney, Atlas, Joormann, Eugène, & Gotlib, 2006). Hence there is no real 'safe' alternative. This is different in our dynamic full-body stimuli: a person walking away can be considered an unmistakably safe situation.

Of course, more differences exist between biological motion stimuli and faces, hence the processes of social cognition involved in interpreting those two stimulus classes may not be comparable. Still, some neural regions seem to be engaged by both, most notably the superior temporal sulcus (Allison, Puce, & McCarthy, 2000). Incidentally, this region has also been found to be more active in social phobic patients than in healthy controls when viewing neutral and emotional faces (Gentili et al., 2008).

Another difference between faces and biological motion stimuli as we used them lies in their potential to induce distinct emotion. Evidently, any emotional relevance attributed to either of the biological motion percepts would be much more subtle than in the explicitly emotional faces used most frequently to investigate this kind of questions. In that context, a recent study that used continuous flash

suppression to render faces invisibly may provide more clues (Stewart et al., 2012). These authors measured the time it took for faces that subtly varied in dominance and trustworthiness to emerge into awareness and found that dominant and untrustworthy faces took significantly longer to break through from suppression than neutral faces. Stewart et al. (2012) recognized that their results were counter to the negativity bias (or vigilance theory), but argued that they may reflect a suppression of a threat-related stimulus, flowing from a passive (avoidant) fear response (analogous to freezing). Interestingly, in a second experiment they asked people to rate their trait submissiveness and propensity to trust using two surveys. Individual personality differences correlated nicely with breaking times in continuous flash suppression. Specifically, for people high in trait submissiveness the breaking times for dominant faces (compared to neutral) was even longer, suggesting more avoidance. People with a high propensity to trust, on the other hand, showed less avoidance of untrustworthy faces (shorter breaking times), compared to neutral faces. Apart from highlighting the importance of individual differences in personality traits for these perceptual biases, these results nicely align with stronger suppression of an approaching percept for people with high anxiety in the current study.

Though Voss et al. (2008) already demonstrated that effects of emotion on perception cannot solely be attributed to response bias, we have a few other reasons to argue that post-perceptual decision bias is unlikely to explain our findings. First, the time gap between screening and the (blind) experimental tests obscured the link between anxiety and the visual task. Debriefings indicated that participants were not aware of the research question, let alone the expected direction of the effect. Second, we had to exclude a similar number of participants from both groups for a lack of variability in responses, suggesting that in both groups participants responded according to their percepts and did not feel obliged to vary their response criterion if their percepts did not change. We cannot completely rule out that other non-specific factors such as general intelligence played a role, but no evidence is available to substantiate such a role. Despite these arguments, future studies should investigate bias in a similar stimulus but lacking the social relevance of our point-light walkers, in order to completely rule out decision bias.

Altered perceptual abilities have been documented before in socially anxious viewers (e. g.better exogenous attention for neutral stimuli, Moriya & Tanno, 2009). This may be at least partially due to their lack of habituation, even to neutral stimuli, as measured by the amygdala and hippocampus activities (Blackford, Allen, Cowan, & Avery, 2012). Further experiments have to clarify to what extent

these more bottom-up processing differences have contributed to our finding. The shorter reaction times in the high social anxiety group for certain levels of perspective information may support this explanation. For example, the strong facing bias in the general population implies that quite a lot of perspective information of a walker facing away from the viewer has to be added for people to actually perceive it so. Highly anxious people seem to pick up these physical perspective cues earlier, perhaps because they are more tuned to these personally relevant situations. In the light of recent evidence that oxytocin administration can improve the detection of biological motion (but not mechanical motion) in noise (Kéri & Benedek, 2009), it is tempting to speculate on a role for this neuromodulator here. It is thought that oxytocin exerts its influence on social behavior through its role in increasing salience of social cues (Bartz, Zaki, Bolger, & Ochsner, 2011). Because levels of endogenous oxytocin increase with symptom severity in social phobia (Hoge, Pollack, Kaufman, Zak, & Simon, 2008), higher oxytocin in our high anxiety group may help them to better process the perspective cues.

A few studies found that looming stimuli (in the visual or the auditory domain) lead to shorter reaction times in (non-anxious) monkeys and humans (e. g., Maier, Neuhoff, Logothetis, & Ghazanfar, 2004; Schiff, Caviness, & Gibson, 1962; Wuerger, Crocker-Buque, & Meyer, 2012), presumably because they are ecologically important signals (analogous to our facing percept). We, on the other hand, found that high anxiety is associated with more receding percepts and shorter reaction times, though only for the lower perspective values (see Fig. 3). Note however, that earlier looming studies mostly used simple expanding disks (or increasing intensity of sounds in the auditory case). In our task, the size of the dots and the full size of the walker remained constant irrespective of perspective level. Whether the stimulus was 'looming' or 'receding' had to be decided based on the more subtle structural dynamics of dot relations. When comparing reaction times for the two different percepts, we found no clear differences, suggesting that both percepts (looming and receding) were formed faster in the high social anxiety group. If, as we proposed earlier, people with high social anxiety are more attuned to the cues in the walkers that distinguish in-depth orientation then that may have caused faster processing and responding. This would mean that the low social anxiety group literally needed more time to settle on an unambiguous percept. Unfortunately, we cannot exclude the possibility that this group did not comply with the instruction (to report initial percept) as well as the high anxious group and hence did not report their actual first percept, explaining the longer reaction times. However, since we did not explicitly

instruct our participants to respond as fast as possible, we do not want to read too much into the reaction time data.

In recent years biological motion stimuli have been fruitfully used to better characterize information processing in people with altered configural or social cognition, such as autism (Nackaerts et al., 2012; Pavlova, 2012). Up till now the conventional task in these experiments is detection of point-light figures in noise. We have shown that processing differences can also be demonstrated in a task probing bistable walking direction. Future research may take advantage of this task and stimuli in two complementary ways. Instead of our emotionally neutral walkers, explicitly threatening body language can be used to elucidate the role of top-down semantic relevance (van Boxtel & Lu, 2012). Structural and kinematic features that are known to influence in-depth orientation can be manipulated to limit the availability of certain bottom-up information (Schouten, Troje, & Verfaillie, 2011).

We explored the role of social anxiety on the perception of a bistable, threat-relevant stimulus and have demonstrated the usefulness of this paradigm for studies investigating differences in social anxiety and more generally influences of (trait) emotion on social perception. Our results indicate that trait social anxiety can bias perception in a way that counters this anxiety. This self-protective bias might be specific to our bistable task with the point-light stimuli, because it opposes the attentional bias towards threat usually found in anxious groups. Future studies will have to incorporate a broader range of social anxiety scores to see if this explanation holds ground and to accommodate the discussed limitations of these findings. The decreased amount of threatening conscious percepts in highly anxious people suggests that trait emotion of the observer and emotional relevance of stimulus material interact in ways unexplainable with traditional models of the perceptual system as disjointed from emotional processes. Our results extend the literature on the influence of emotion on perception, showing that not only emotional states but also traits can bias perception, though not necessarily in a detrimental way.

Happiness is neither virtue nor pleasure nor this thing nor that, but simply growth. We are happy when we are growing.

— W.B. Yeats

6

The place of value in the predictive brain

From 1932 to 1934 Wolfgang Köhler, one of the great Gestalt psychologists, gave a series of lectures with the title "The place of value in a world of facts" (later turned into a book, Köhler, 1938). In it he discusses the alienation people often feel when confronted with science, which casts a world indifferent to the rich meanings and values they ineluctably experience it through. How could they ever feel at home in a universe that only has room for blind facts and forces? Although their answers left a lot to be desired, Gestaltists always had a knack for posing fascinating and important questions. Somewhat updated to the current day, the issue becomes how value and meanings enter a world of matter and information. Since Köhler's lectures, Shannon's theory of information and Turing's theory of computation have revolutionized psychology, because their pioneering work suggested one could talk about information processing in the mind in a well-specified and quantifiable manner. Still, the gap that Köhler described is not bridged; Shannon's information is not meaning —it does not have intrinsic relevance or value. Maybe the information processing paradigm, inspired by the computer metaphor, was bound to find its limits when confronting issues of value and emotional relevance, that seem so grounded in our existence as biological creatures. However, the perceptual system, as the interface between organism and world, is a good starting point for an inquiry into these matters, as Gestalt psychologist also realized. The perceptual system is where tractable differences picked up from the world will need to be turned into relevant meaning for the organism. However, we will need a theory that is not limited to perception, but has the potential to capture the key characteristics of emotional value.

More than 130 years after psychology's secession from philosophy, we do not have a standard model of cognition (including emotion; the dissociation has always been contrived). This may very well be one of the reasons for the current methodological crisis in psychology. Newell (1973) lamented that we as psychologists never

put things together, we just juxtapose intricate but disparate experiments. This problem has only deepened since, due to the crave for flashy, ad hoc and counter-intuitive findings in the current publication climate. On top of that, unification and exchange of knowledge is hampered by brand names in psychology (Taatgen & Anderson, 2010). Research groups try to hegemonically push their terminology, obscuring much of the fruitful commonalities among them, and sometimes resulting in duplicate work. To be fair, this is not necessarily an intentional strategy: with booming publication numbers, it is becoming increasingly difficult for psychologists to find the time to read and critically evaluate the work of others outside their specialized field, let alone integrate it in their own work.

One attempt at unification that has in recent years been shown to hold a lot of promise is predictive coding and its core principle of Prediction Error Minimization (PEM). It is likely the first theory that naturally blends perception, action and cognitive beliefs into a coherent, well-founded framework, pleasingly taking down the walls between these subfields. Although plenty of fundamental issues concerning its computational articulation and biological implementation remain (see commentaries on Clark, 2013b), as a unified theory of cognition it arguably fares better than any other alternative we have. But it is an ambitious approach, proposing that PEM is all the brain does. And of course, if it is to become an overarching framework of the mind-brain, emotions somehow have to fit in. After all, a general theory of the mind-brain that does not address emotion, so central to existence and survival, is painfully incomplete, even if the experiential aspects of emotion would turn out to be purely epiphenomenal for the account. So far, there is very little work on affective relevance and experience in predictive coding. One main reason for this is that the unifying logic of predictive coding (a single computational principle for the whole brain) is directly opposed to the popular notion in emotion theorizing that emotions are a bricolage of modules adapted to very specific challenges in our ancestral environment. Rather than built around a single neat, optimal logic, emotions are assumed to be a messy, ad hoc bag-of-tricks. However, in practice, it has proven difficult to distinguish different emotion 'modules' in the brain, even at the subcortical level, which has led some emotion theorists to abandon this route in favor of a view that assumes fewer fundamental affective 'building blocks'. The aim of the current chapter is to show that this movement may afford new ways to integrate emotion in predictive coding. Much of this is, as we will see, thanks to the clear evolutionary rationale that is at the core of predictive coding and its PEM.

At first blush, this may seem to lead to a concept of value or emotion that seems rather alien or counter-intuitive to how we usually think about emotions. However, it will turn out to have much in common with existing theories of emotions, such as the meta-monitoring account of emotion and Carver & Scheier's (1990) control theoretic model of emotion, appraisal theory and conceptual act theory. Most importantly, perception (information processing) and emotion will be shown to be entangled from the start.

6.1 GENERALIZED PREDICTIVE CODING

That expectation is inextricably bound to perception, often only becomes apparent when expectations go wrong. A colleague changed her haircut. Someone took away the office plant. These instances show that expectations are constantly there coloring our perception, whether we are explicitly aware of this —as in a melody, where expectations extend perception beyond the given note— or not. This predictive transcendence of given input is what furnishes Gestalts, whether it be temporal prediction in melodies or spatial prediction in faces, where one location reliably predicts input in another.

It might therefore seem somewhat surprising that it took till recent years for models of perception to appear in which prediction takes front and center. To be fair, the importance of expectation in perception was acknowledged as far back as Saint Augustine and William James. But the idea of a passive —and 'objective'!- perceiver that could absorb raw sense data just remained very appealing in perception science. With Bayesian accounts of perception, gradually prior knowledge was given an important, biasing role but arguably only with predictive coding, the active, constructive role of the perceiving organism is starting to be fully recognized.

Predictive coding holds that an organism is constantly, proactively predicting the inputs from its environment. Since it has no independent access to objective features in the world, all an organism can do is learn patterns in its input generated by statistical regularities in the environment and by its own actions (Clark, 2013b). While in principle there may be different ways in which prediction could modulate perceptual processing, predictive coding proposes a well-defined computational scheme and a single guiding principle. The scheme describes that every level of the perceptual hierarchy predicts activity in the level below, in effect explaining away input that is consistent with it such that only mismatches remain. These mismatches, called prediction errors, are sent upwards to update future top-down predictions. Predictive coding is not in the first place a theory about temporal anticipation (the colloquial sense of prediction) but rather about signal prediction: one region of the

brain "predicts" what quasi-simultaneously happens in another region. As Niels Bohr reportedly remarked, prediction is very difficult, especially if it is about the future. Hence, much of the brain's predictive activity has a limited time frame. However, the higher up in the hierarchy, the more time and space predictions can span, because they can work with regularities defined on lower levels. In this way lower level predictions model the faster changing dynamics, while those higher up track and recreate slower changing dynamics.

Predictive coding thus completely inverts the classical bottom-up view of the perceptual hierarchy. Top-down influences do not modulate the main bottom-up stream. Rather, in line with the 'controlled hallucination' view of perception (Horn, 1980), the brain actively generates the perceptual world (predictions are based on generative models, i. e., models that can generate the input), and perceptual input is in fact the feedback on how good these constructed models are. As such, it is more appropriate to say that bottom-up information modulates the intrinsic dynamics of the organism. Although anatomically prediction errors are conveyed by feedforward connections, functionally they are the feedback, sanctioning the models we construct of the outside world.

The fundamental underlying principle guiding this process of iterative, hierarchical matching of predictions with inputs is that of prediction error minimization (PEM). Perception is inference to the best prediction, the one that minimizes prediction errors. Simultaneously, learning will use remaining prediction errors to home in on the best predictions for the current context, thereby reducing future prediction errors. Hence, we perceive what led to successful predictions in the past (see also Purves, Wojtach, & Lotto, 2011). There is one other, complementary way of minimizing prediction errors, which does not focus on improving predictions, but rather on modifying the things predicted. We are of course talking about action here. In this framework, movements serve to bring the input closer to our prior expectations. More specifically, they are induced by their expected exteroceptive and proprioceptive consequences (Friston et al., 2010), much in line with James' "anticipatory image" (James, 1890) and with the ideomotor principle (Hoffmann, 2003). Like object-level, conceptual predictions ("an apple") unpack to a myriad of lower-level featural predictions ("green", "curved",...), so can high-level expected states ("goals") be unpacked in specific component predictions and eventually in expected proprioceptive states (acting as motor commands). When the latter are compared to afferent signals of muscle spindles at the spinal level, they generate sensory prediction errors to be reduced by motor neuron activation, in a classical reflex arc. Hence, motor commands are replaced by

expectations about the state of proprioceptive sensors. At a higher level these 'commands' stem from beliefs about state transitions. A certain perceptual stimulus may be predictive of a state transition through the agent's intervention (an affordance, if you will), that can be actualized by unpacking this prediction to proprioceptive states.

Bear in mind that, from this inference system's perspective, there is no intrinsic difference between the external and the internal milieu. With the same predictive machinery, generative models can be learned about changes in interoception, based on input from somatovisceral sensors (Seth, 2013). Likewise, internal 'actions', such as autonomic responses, are brought about by similar principles as 'external' actions. They consist of changing a bodily set-point or expectation (e. g., temperature) so autonomic reflexes (e. g., shivering) can be elicited. As a concrete example, a stimulus that is predictive of an imminent drop in temperature, may induce such anticipatory shivering by changing the expected body temperature, in order for the ultimate prediction error, when the actual cold sets in, to be smaller or short-lived (more on the link with homeostasis below).

In sum, this is how perception and action can be subsumed under a common principle of surprise minimization. But note that all this concerns surprise (prediction error) in the informational sense, also called *surprisal*, which is different from agent-level surprise (in the common experiential sense). A very surprising interpretation for an agent ("an elephant on a stage") can still be the prediction that best explains current input (reduces surprisal), however unlikely the occurrence of such an event (Clark, 2013b).

So far we have not covered the whole predictive coding story. To see why it is still incomplete, notice that an agent has no independent way to disentangle noise, i. e., variation that is unlikely to repeat in next instances, from signal, i. e., actual learnable changes in the regularities in the environment (J. Feldman, 2013). In other words, not all prediction errors are created equal. To properly update models only when needed, errors at every level have to be weighted according to their expected precisions (inverse variance). Predictive coding without precision is as problematic as comparing means (prediction error) without considering variance (inverse precision) in statistics. Precision makes sure reliable differences are learned, while noisy differences can be ignored, to avoid loss of future predictive power (cf. overfitting). Notice that expected precisions are also a form of predictions (meta- or second-order predictions, predictions about prediction errors) that have to be learned from the same input as the 'first-order' predictions, using the same error minimization mechanisms. Indeed, levels of noise also vary predictably across different contexts, and these regularities should

be learned to differentially weight prediction errors. This is particularly crucial in volatile environments where reducible (learnable) uncertainty linked to actual fluctuations in the predictive structure of the environment, coincides with irreducible uncertainty due to probabilistic noise. The precision mechanism of predictive coding should therefore be conceived as a meta-learning mechanism, for learning what is learnable (Gottlieb, 2012), or equivalently, determining the informative value of prediction errors (Behrens et al., 2007). This suggests that what is commonly understood as attention should be considered as the result of precision tuning. Turning up the precision (high weight) of prediction errors at a certain level means that top-down predictions higher up will not be successful in explaining input (i. e., will not survive), so matching will "zoom in" on the level in question. This seems a plausible way in which attention is deployed, at the same time regulating the relative influence of bottom-up information flows (prediction errors) versus top-down predictions in perceptual inference. In a related way, precision is also involved in action. To initiate an action, be it internal or external, prediction errors concerning the current state have to be attenuated. If not, they will update predictions higher up, rather than induce action (Adams et al., 2013; Seth, 2013). By lowering their precision, their impact is transiently reduced to enable action. The neural mechanism thought to be responsible for precision setting is gain modulation (by neuromodulators) on the output of units communicating the prediction errors (H. Feldman & Friston, 2010; Vossel et al., 2014). Finally, note that precision constitutes another way of PEM, namely by effectively silencing remaining prediction errors —making them powerless by reducing their gain.

We limit ourselves to this brief sketch of predictive coding and refer to the many in-depth resources for more details about its computational mechanisms and how these map onto neural circuits and their plasticity (e. g., Bastos et al., 2012; Friston, 2003, 2010a). Further implications of the framework will be discussed to the extent that they connect to value and emotional relevance.

6.1.1 *Re(de)fining information*

Although the cognitive sciences intend to formulate information processing theories of the mind (or one of its faculties), the concept of information itself is insufficiently questioned. One of the reasons is a tacit assumption that Shannon adequately and sufficiently described information. It is beyond doubt that Shannon's theory of information was of tremendous importance for cognitive science (not to mention technological advancement). Arguably the

success is in large part due to the fact that he was able to reduce information to a concept that could be fully formalized (Deacon, 2010). The big accomplishment of Shannon is that he realized that any physical difference could be information, so that this capacity was independent from the material carrier. However, Shannon's information is not about something, so it may be better characterized as information-bearing capacity rather than actual information (Deacon, 2010). Meaning (plainly 'information' as commonly understood) does not come into it. There are receivers but no interpreters.

So what else is needed to approach the full sense of information as meaningful content? Gregory Bateson (1979), one of the fathers of cybernetics, gives an important clue. He described information as a difference that makes a difference. To create a difference, he reasoned, one needs at least two versions of something. Bateson's example is binocular vision: two eyes hold two versions of input that, when compared, provide new information on depth. In line with predictive coding, one can say that the brain also has two versions, one that it generated itself (the prediction), and one that is provided by the sensory input. The difference (or prediction error) that is made possible by these two versions, may become new information. Note that this explains why an absence or non-occurring of something can be information, when the other version predicts something that does not happen.

However, the second part of Bateson's definition required this difference to matter or "make a difference". In perception theories this point is rarely acknowledged. They often assume that any physical difference (within the range and resolution of the sensors) is captured and thus is information for the visual system. This is grounded in a naive correspondence theory of perception, which says that properties in the world have simple mappings to perceptual experience. The crucial insight of the more recent interface theory of perception was that a difference might not make a difference at all (Hoffman, 2009; Koenderink, 2010). The interface theory argues that perception, as an evolved capacity, does not need to be literally truthful to allow adaptively efficient functioning. Indeed, a non-veridical perceptual system may be more cost-efficient. Perception has to provide a kind of user interface, similar to the desktop-with-icons interface that enables easy interaction with a computer without being a veridical representation of its underlying workings.

Since efficient interaction with the environment (for survival) is the criterion for evolutionary selection, agents with a good user interface may actually drive agents with veridical perception to extinction, provided that perceptual processing has some non-negligible cost (Mark, Marion, & Hoffman, 2010). Encoding all physical (veridi-

cal) differences is wasteful, because some of them do not make a difference in guiding adaptive behavior. As a very simple example, say there are three different signals (s1, s2, s3) but only two relevant actions, a1 for s1, a2 for s2 *and* s3. It would be wasteful for the agent with only this signal-action repertoire to be able to distinguish between s2 and s3 (see simulations in Mark et al., 2010 for more elaborate examples). The implications of this with regard to animal perception are easily accepted (and described with von Uexküll's concept of *Umwelt*) but that human perception might be subject to similar constraints is often dismissed. Bateson appreciated that the usefulness of a representation of reality does not necessarily require exact veridicality, but only some preserved relations with regard to the purpose at hand. More recently Bitbol (2007) emphasizes: "Success of an action does not demonstrate the truth of the assumptions under which this action is performed; it only demonstrates [...] their efficiency as guiding principles."

In the above we have relied on your intuitive grasp of what it means to 'make a difference'. We used concepts such as 'relevant' action, usefulness, purpose, disregarding more mundane types of making a difference. Moving a stone in a river is a difference that changes (makes a difference for) the course of the water, but we would not call that information. Arguably, making a difference implies more than having this kind of consequences. The challenge is to describe "making a difference" without recourse to a kind of teleological homunculus. It seems that we can only truly speak of information in the context of a system that actively picks up differences that can impact on its own functioning. Information seems to be bound to systems that are somehow prepared for this information. This can be illustrated with Koenderink's (2010) 'Sherlock model' of perception. A particular question will make some difference into information. The mud on the shoes will suddenly become information, relative to Sherlock's plot (prediction). Under a given prediction, a difference can emerge that makes a difference in the sense that it confirms or disproves the prediction. Perception means probing input with predictions, forms of regularities already present in the perceptual apparatus.

So, it appears that predictive coding is all about finding differences that make a difference —about creating information as relevant differences. Relevant prediction errors are those prediction errors that lead to different conclusions within the current context, namely, require an update of existing predictions. In a minimal sense this is "making a difference". A central instrument to accomplish this is precision estimation. It makes sure that any difference (prediction error) is not taken at face value but weighted by ex-

pected noise of the situation. This is an estimation of whether it should make a difference on the system's predictions. Although it is easier to conceptualize precision estimation on lower levels than higher up (i. e., it is clearer what noise is), it is assumed to work on every level of the perceptual hierarchy. Hence, what is irrelevant difference is also relative to the level. A typing error is an irrelevant difference when figuring out the meaning of a text, but it can become information when judging personality of the writer, e. g., diligence (Deacon, 2010).

Importantly, as in interface theory, relevance or usefulness has to take into account both costs and benefits (Sperber, 2005). Predictive coding indeed constitutes a form of sparseness of representational resources. For one thing, predictability implies compressibility. Precision estimation further aims to dampen differences that do not matter, saving resources for those prediction errors that likely need to update the predictions (i. e., are more than just noise with regard to the question asked). This is a continuous, fallible optimization problem, which may imply that in case of low system resources, prediction errors might be ignored that otherwise could actually have had important implications. Alternatively, in case one subsystem has high expected precisions (demands resources), differences in different subsystem may be ignored, leading to a form of inattentional blindness.

6.1.2 *Prediction and homeostasis*

We described predictive information as "making a difference" in a minimal sense, without referring to the organism that uses this predictive sensory system. It is clear that non-predictive information is useless to the organism (Bialek et al., 2001). This does not mean that all predictive information is useful (~makes a difference), but predictiveness is at least a lower bound on relevance. What *is* relevance in the context of an organism?

Organisms have to 'pay' for their own organization (order) by increasing the entropy of the wider system in which they are embedded (cf. second law of thermodynamics). Bioenergetic regulation by homeostasis is how they accomplish this and hence it is the most fundamental goal of any organism. Of course, this is not an intentional goal. It is just a result of the fact that organisms that do not tend to homeostasis will lose existence as a unit. One should view homeostasis as a set of expected states installed by evolution, because they have proven to enable continued existence.

Importantly, homeostasis is not actually static. Organisms do more than reflexively respond to perturbations with compensatory ac-

tions (e. g., expressing those genes that will facilitate continued existence in the new circumstances). If reliable predictive information is present, it is more effective to *anticipate* changes with compensatory action. Indeed, this is precisely what even simple organisms do (Freddolino & Tavazoie, 2012). For example, when the single-celled gut bacterium E. coli is ingested by mammals it will respond to the temperature shift by not only upregulating heat shock genes (to compensate for temperature) but also by downregulating genes for aerobic respiration. They use the temperature information (when entering the mouth) to predict that they will end up in a low oxygen environment (gastrointestinal tract). This is quite an astonishing feature for a micro-organism. As Freddolino and Tavazoie (2012, p. 369, our emphasis) describes: "microbial behaviors are as much responses to the *meaning* of a specific environmental perturbation (viewed in the context of their evolved habitat) as they are responses to the direct consequences of that perturbation".

Clearly, prediction emerges from the fundamental need to compensate for environmental influences disturbing the internal milieu and therefore threatening further subsistence of the organism (Sterling, 2012). Predictive processing subsumes reflexive homeostatic regulation, because the latter is a prediction that a certain environmental property is stable enough to merit an immediate response (Freddolino & Tavazoie, 2012). In general, however, reactive compensations will be less efficient than predictive ones (Heylighen & Joslyn, 2001).

Predictive homeostatic regulation requires reliable spatiotemporal regularities in the environment. For E. Coli the predictive "learning" of these regularities does not take place within the organism but within populations. These populations respond to milieus with more uncertainty by diversifying (gene expression noise). Those organisms that then capture a new (stable) regularity gain fitness (increased reproduction). The predictive environmental relation is now embodied in the molecular regulatory networks of the cell (Freddolino & Tavazoie, 2012). Environmental regularities left their imprint on the organism's constitution, just because a constitution embodying these regularities increases fitness in a Darwinian sense. In analogy to predictive coding, evolution can be considered as an error-correcting code, except that the errors are not represented on the level of a single organism. However, a good match between the organism's predictions and its environment will lead to better survival. Still, the normative character (the value or relevance) originates in the organism, not in the process of evolution (Deacon, 2011). Certain predictions support the continued existence of the organism. This applies to plants, that, for example, embody sea-

sonal shifts, as well as to animals that by their physical constitution are adapted to their environmental regularities (Rosen, 2012[1985]). In animals, "the gross bodily form, biomechanics, and gross initial neural architecture of the agent all form part of the (initial) 'model' and [...] this model is further tuned by learning and experience" (Friston, Thornton, & Clark, 2012).

What distinguishes simpler organisms from more complex ones is then the plasticity (flexibility) to build new predictive relations within the lifetime of an organism (a.k.a. learning). For example, fixed action patterns described in ethology are predictions that set in motion stereotyped behaviors given some innate releaser stimulus (a more elaborate version of we saw in E. Coli). Little if any flexibility is possible here, as is almost comically clear from the ethologists' examples (e. g., the egg-retrieval behavior of the Graylag Goose that works as good or even better on vaguely similar objects, like door knobs; Tinbergen, 1951). On the other hand, conditioning can be conceptualized as *flexible* learning of predictions (McNally et al., 2011; Sutton & Barto, 1981); Pavlov's salivating dog predicts that food will follow ringing of the bell. Or when the autonomic nervous system predicts and anticipates the energetic demand of upcoming exercise based on environmental or internal cues.

Let us, as an example, briefly consider a prediction of intermediate flexibility, as it appears in the circadian clock system. The day-night rhythm is a steady regularity that the bodily clock system is tuned to, mainly coordinated by the subcortical suprachiasmatic nucleus. The problems associated with jet lag and shift work, show that this system is not prepared to deal with large, abrupt shifts in light-dark cycles. It evolved at a time in which only slow shifts, specifically the seasonal changes in light exposure, were present (Karatsoreos & McEwen, 2011). Although the clock system can be resynced by exteroceptive light cues (and other interoceptive "Zeitgebers"), it maintains a strong endogenously powered prediction, even without any light cues (Reinberg & Smolensky, 1983). Errors with regard to this prediction (unpredicted light levels), seem unable to efficiently update the regularity generator (i. e., the prediction), as assumed in predictive coding. It shows that even in the (subcortical) brain of complex mammals inflexible expectations exist, albeit close to the bodily systems (the clock coordinates energy regulation in the body; Sterling, 2012).

To exist, a system has to constrain the number of states it can be in. This innate, homeostatic set of expected states is honed by evolution, through its contribution to survival. Specifically, the system has to keep values of a number of so-called essential variables in check, by error-correcting. These are the variables that evolution 'discovered'

to be important to monitor (Phillips, 2012). When adaptive, this will be expanded to monitoring 'auxiliary', non-essential variables, that are predictive of disturbances of essential variables (Heylighen & Joslyn, 2001). As we see in higher mammals, and especially humans, this may ultimately lead to complex hierarchies of learned predictions (generative models), to account for a changeable environment. The evolution of long-range, high acuity visual capacity in mammals presumably had a major contribution to the expanded predictive capacities of the mammal brain. It literally created room for elaborate prediction and planning, rather than merely reactive behavior (MacIver, 2009). But most importantly, the fact that the relevance of environmental features is very changeable within an organism's lifetime[1], means that evolution fell short in tracking regularities and flexible predictive capacity had to compensate. PEM will make sure that the agent will capture the statistical regularities in its environment and hence become a model of this environment, in line with Ashby's Good Regulator theorem: "every Good Regulator of a system must be a model of that system." (Conant & Ashby, 1970).

Predictive coding can be such an encompassing theory because it is rooted in control theory from cybernetics (e. g., Powers, 1973). Control systems make adaptive behavior possible through an open interaction with their environment. They invariably accomplish this by setting an expected state and feeding back the input-induced deviance to correct the internal dynamics. Any living organism consists of multiple hierarchical predictive control layers. Bechtel (2009) explains:

> "When we conceptualize control, we often think hierarchically and situate all decision making at the top of the hierarchy. This, however, works poorly both in biology and social institutions. As a result, biological systems usually have multiple layers of control arranged such that higher-level control systems can bias the functioning of lower-level ones (often by affecting the conditions under which more local control systems operate), but do not directly determine the behavior of the lower level systems. This can be appreciated by focusing on organisms in which cortical level control systems have been removed —in such cases many functions continue unimpaired but cannot be (directly) coordinated in the service of higher level objectives." (Bechtel, 2009, p. 13)

1 Part of this may have had to do with the social world.

Built-in prediction or priors of organisms can go beyond homeostatic reference values or other constraints imposed by bodily constitution (Scholl, 2005). For example, the frog's nervous system comes prepared to see moving blobs of a certain size (to catch flies) (Lettvin, Maturana, McCulloch, & Pitts, 1959). It expects to see these blobs, because if it wouldn't it would dramatically jeopardize its survival chances. Similarly, a fish expecting to sense dry land is heading for a certain death. In short, organisms come equipped with sets of constraints: priors or predictions about states that the organism needs to frequent to be viable. Survival "depends upon avoiding surprising encounters and physiological states that are uncharacteristic of a given phenotype" (Friston, 2009).

Although it clearly evolved for its survival benefits (making sure that the agent revisits a limited set of states), the predictive coding scheme can discover and recreate any statistical structure, given the necessary resources, such as provided by a great cortical expansion. Just specify the relevant variables and a scheme such as PEM can be used to mine inputs for new variables that predictively relate to them (Phillips, 2012). This explains why it is particularly powerful and versatile in humans, where it is harnessed to support an endless variety of goals, concerns, expectations and predictions. But although it gains some independence from evolved variables, it is important to keep in mind that, even in humans, it is still ultimately grounded in and bounded by homeostatic, bodily constraints. Hence all distinctions we can make originate in those constraints. These are constraints (predictions) that allow the creation of new constraints (predictions) (Deacon, 2011). We use prediction errors to differentiate and build multiple levels of predictions. Crucially, the boundedness in the homeostatic set also gives the whole predictive endeavor its normativity (the 'mother-value of all values'; Weber & Varela, 2002). Once the organism engages itself to make a prediction, however insignificant, there is something at stake, because of the link from the quality of predictions to basic organismic functioning. There is a vested interest for the prediction to materialize. Mental models are attached to bodily existence. Reducing prediction error can be a proxy for fitness, but this by no means implies that only predictions that improve fitness can be formed. For example, even though some perceptual predictions may seem not to be consistent with expected states, it is all the more important to make them accurately (and not hallucinate more agreeable alternatives).

Importantly, if high-level objects or concepts (e. g., about social relations) are more predictive of the essential variables than lower level input characteristics, then the capacity to form these will evolve.

As such, Gestalts originated as functional, high-level representations that are actually most predictive of fluctuations in homeostasis. But once the predictive machinery is in place, these (Gestalt) predictions have independence from the concerns the predictive capacity itself evolved for, because the PEM principle is a purely informational criterion.

Aside from the differences in abstraction (and those in flexibility described earlier), there are several important transitions to discern in evolution regarding the kinds of predictions that can be generated. Phillips (2012) outlines: "from predictions only of things that are directly observable to estimates of things not directly observable; from generative models averaged over various contexts to those that are context specific [...] from hypothesis testing to pure hypothesizing freed from testing." (Phillips, 2012, p. 12). The latter of course opens the way to a form of "offline", simulated predictive coding (see below) necessary for complex planning and thought. Each of these transitions deserves to be spelt out in more detail, but that is beyond the current objectives.

We can envision a historical path dependence (similar to evolution) for PEM within a human's lifetime. Any human brain will construct a complex predictive mental edifice, unique in its intricate branches. This is in a very real sense you, more so than the material you. The different expected sensorimotor regions that act as attractors, will vary dramatically within and between individuals, provided that on the whole a small set of physiological states will be frequently revisited. PEM is the proximal, local mechanism that makes sure that in the long run organisms stay within physiological bounds (Friston, 2010a).

By explicating these links between homeostasis and generalized predictive coding, we have set the stage for emotional value to be introduced.

6.2 DRIVES, REWARDS, AND AFFECTIVE VALENCE

The important insight, entailed by generalized predictive coding, is that any "drive" one can describe for an organism, is better specified as error reduction relative to an expected reference state (prediction) (Keramati & Gutkin, 2011). A drive or motivation is then determined by the discrepancy between current and expected state. An organism will mobilize resources to reduce this prediction error, but because these predictions are often less flexible than learned perceptual predictions, the errors are mostly reduced, not by revising predictions, but by behavior, i. e., performing actions to change sampling of the world (see above).

In this scheme, rewards are just expected or familiar sensory states (Friston, Shiner, et al., 2012). Intuitively we feel that we avoid punishment or seek reward and therefore visit these states less or more frequently, respectively. Predictive coding turns this intuition upside down, describing frequently visited states as rewards because they are expected. Rewards do not 'attract behavior', but attainment of rewards is the result of prediction error minimization, exactly as described for perception and action in general.

As we saw, the repertoire of innate expected states is specified and extended by learning throughout an animal's life. In fact, within this view, there are no distinct reward or punishment stimuli (Friston, Daunizeau, & Kiebel, 2009; Wörgötter & Porr, 2005). Any sensory signal has a cost, namely the prediction error. It tells something about the success (failure) of the generative model we used for predicting the input. Given that only relevant predictions are learned, the short-term, moment-to-moment effort to reduce prediction errors translates into long-term goal of survival.

This implies that habits or 'rituals', i. e., predictable sequences of behavior, are in fact rewards. Animals, including humans, are indeed creatures of habit (James, 1890; Darwin, 2011). There is usually no tremendous pleasurable experience to habits (we will get back to this point later on), but not performing habits when eliciting context is present seems to produce some negative affect, indicative of some error (drive). It speaks to the self-sustaining nature of habits (Egbert & Barandiaran, 2014; Egbert & Canamero, 2014). Indeed, for over-learned fixed behavioral patterns, devaluation of the reinforcer that was originally used to establish the behavior will not lead to reduction in behavior (Wood & Neal, 2007). The wider implication is that organisms do not only preserve their life (homeostatic predictions) but also their *way of life*, as a set of expected (preferred) behaviors (Di Paolo, 2003).

This conception of reward fully supports the idea that rewards and punishments are always subjective and intrinsic, meaning they "are constructions of the subject rather than products of the environment" (Dayan, 2012a, p. 1089). They are dependent on the drive state or the position relative to other expected states. Reward is not something in the environment, much less an external critic such as often assumed in computational reinforcement learning (see a similar critique in Singh, Lewis, & Barto, 2009). But psychological theories too, often incorporate a semi-hidden homunculus. Here, the value (or cost) function applied to perceptual or cognitive output hides an 'evaluator', an unanalyzed 'agent' that can assign the values, within an allegedly objective, quantifiable construct. Misled by our intuition that rewards are self-evident, these homunculus rem-

nants too often go unquestioned. Another reason why we might want to revise the concept of reward, is to circumvent its inherent circularity. Reward is often defined as a stimulus that elicits (reward seeking) behavior (Schultz, 2007). Evidently, one cannot invoke rewards to explain that same behavior later on (Friston, Shiner, et al., 2012).

As a rule, it is possible to reproduce any reward/cost function with appropriate predictions (Friston, Shiner, et al., 2012). So, rather than running the outcome of perceptual inference through a reward/cost function in an additional step, inference and value are always entangled in generalized predictive coding. This is an appealing property, from the point of view that emotion is inextricably bound with perception and action (Barrett & Bar, 2009). The quality of our interaction with our environment is evaluated continuously, using predictive testing of our mental (and bodily) models in perception and action. Hence, predictive processing incorporates cost/benefit computations —prediction errors directly inform us about our success in dealing with the world.

This different framing of reward does not mean that learning or behavior is not as constrained as in conventional models of reward and punishment. To take the extreme example, even if at the agent-level a pain stimulus is perfectly expected, across all levels of predictions this will never become an expected state. Tissue damage can be seen as a violation of a bodily expected state (bodily integrity) that is not compatible with continued existence. On the other hand, this approach has no difficulty explaining why humans seem to find reward in endlessly varying individual "niches", based on the wide flexibility in predictions they can generate. Humans can derive rewards from arbitrary, idiosyncratic stimuli, and probably evolved this way because utility of stimuli was very changeable (often socially defined, internalized expectations). Other animals are more tied to particular ingrained expected states, but also show species-specific innate 'reward-niches'.

Two other developments started eating away at the concept of value as absolute, stable representation of utility, guiding decision making. First, Chater and Vlaev (2011) convincingly argue that, similar to sensory judgment in psychophysics, value is a not represented as an absolute magnitude but rather as a comparison, relative to the local context. They garner a wealth of evidence showing that value judgments are extremely context sensitive and malleable (e. g., through the set of options, preceding items, etc.). Neural data as well do not support the idea of value as a stable, common currency, but rather point to a relative encoding dependent on the local, momentary perceptual and cognitive context. Such a comparative view,

where rewards are not defined independent of the specific stimuli and the perceptual-cognitive context, is much more consistent with generalized predictive coding. Chater & Vlaev conclude that "to the extent that people have a grasp of their own, more global, values, this must be inferred from sampling their own past choices and other memories, thus revealing their preferences" (Chater & Vlaev, 2011, p. 96). In other words, humans can easily infer the reward value based on experience sampling, but these values are *constructed* predictions that best explain the sampled experiences. We are able to develop abstract representations (predictions) of reward that may start to play a role at another, possibly conscious, level of reducing prediction errors. Generally however, these representations are not necessary to enable adaptive behavior.

The second development exemplifying the move away from a value-based approach is the road taken by Solway and Botvinick (2012) and Schwartenbeck, FitzGerald, Mathys, Dolan, and Friston (2014). These researchers reframe goal-directed decision making as Bayesian inference instead of reinforcement learning. Rather than finding the policy that maximizes expected reward, this approach assumes reward attainment and finds the policy (state-action pair) that best explains or causes that effect (abductive inference, same as perceptual inference). If one redescribes utility of outcomes as prior beliefs about states one will end up in, one can use the same predictive coding machinery to minimize errors along the road to the expected state. This boils down to building a generative model of reward. Importantly, it requires that we have prior beliefs about what the world will be like *and* about expected final states (goals) (Moutoussis, Fearon, El-Deredy, Dolan, & Friston, 2014). The latter are the alternative outcomes that we expect to be reachable with policies we can apply. The key is to reduce the discrepancy between the likely and the expected outcomes. Hence, we can use the inferential scheme of predictive coding for goal-directed learning and decision making, instead of the computationally intractable traditional reinforcement learning schemes (Solway & Botvinick, 2012).

Based on these developments, we conclude, together with Gershman and Daw (2012), that the strict segregation of probabilities (perceptual processing) and utilities (reward processing) is untenable. For one thing, "increasing sensory uncertainty in a rewarded sensory decision-making task decreases expected reward" (D. R. Bach & Dolan, 2012, p. 583). Indeed, reward-modulated activity is found throughout the visual hierarchy, including the primary visual cortex and the lateral geniculate nucleus (Gershman & Daw, 2012; Serences, 2008). Additionally, the entanglement of cost-reward and perceptual processing is clear, if one wants to take into account the

cost of processing or representation. There is an obvious trade-off between the utility of being more accurate, and the cost of computation and representation of those more elaborate models (Gershman & Daw, 2012). Such optimizations are at the core of predictive coding, with expected uncertainties (precisions) determining whether it is useful to capture more differences in input (by more specific predictions).

Intuitively, the identification of prior probabilities (predictions) with utilities seems wrong-headed, for example, Gershman and Daw (2012, p.306) ask: "Should a person immersed in the 'statistical bath' of poverty her entire life refuse a winning lottery ticket, as this would necessitate transitioning from a state of high equilibrium probability to a rare one?" To start to defuse this argument, one has to acknowledge that in such complex cases there is not just one prediction (e. g., of poverty) at play, but rather a complete predictive hierarchy (most of them not even conscious). This person growing up in poverty does not lose his or her expectation to be well-fed and to provide for kin. On top of that, there might be interiorized social expectations, that could also urge the person to accept the winning lot. That said, once accepted, the new situation (the new inputs) may create quite some prediction errors (i. e., unrewarding) given a predictive system unadapted to that new state of affairs (indeed, most lottery winners like to continue their life, including job, as before; H. R. Kaplan, 1987).

Later sections will hopefully shed a different light on these forms of 'upward mobility' (Gershman & Daw, 2012). In general, more work is needed on how we distinguish likely and desired states if both are ultimately predictions. In the end, however, the answer seems to be in the hierarchy, with likely states being about what happens when I see (or do) this. In contrast, desired states are about what I can, more abstractly, expect given my experience and (sensorimotor) capacities (i. e., predictions). At least in humans this seems to have an important social comparative component as well: our predictions are formed based on what people that one considers to be similar to oneself, could attain. The lower level, likely states mostly pertain to faster changing dynamics in inputs (regularities in shorter time frames). The higher level desired states link to slower changing dynamics (e. g., 'I am a good person'). If evidence mounts that undermines such predictions, a full-blown existential crisis may occur. Luckily, there are ways to shield such prediction errors, i. e., to explain them away with 'auxiliary predictions'.

6.2.1 *Conflict theories of emotion*

At first glance, a prediction error account of emotion seems to have much in common with a classical group of emotion theories, called conflict or discrepancy theories (Festinger, 1962; Hebb, 1946; Mandler, 2003), more recently defended by Phaf and Rotteveel (2012). These hold that emotions are caused by a conflict (dissonance) or match (consonance) between goals or other representations and actions or percepts. Negative emotions are often assumed to originate from expectancy violation, which could be constructed as a kind of prediction error. However, remember that predictions are broader than expectations as commonly understood (e. g., see temporal predictions vs. signal predictions), but the idea of dissonance as it pertains to two concepts held simultaneously in mind is broader and seems more akin to prediction errors: one thought has implications that run counter (implications of) other cognitive or action representations.

In these theories, the expectations generally concern high-level conceptual states of affairs or goals. While prediction errors are local, ubiquitous computational products, the violations in these theories are more abstract and conceptual. Alternatively, they concern actions that do not have the specific expected effects. Note that a prediction error on a higher level ("it's not a plane, it's batman") likely has more widespread implications for interpretations and errors in lower levels. But there is another thing that distinguishes these expectations from predictions (in PEM). We only speak of expectation violations when our brain actually made a notable prediction, i. e., with high precision or confidence. Indeed, predictions differ in strength. Compare, for instance, predictions we have with regard to human beings about the fact that they cannot fly, that they do not have feathers (strongest predictions sound most silly if one actually articulates them), versus predictions about incidental features (those that do not define membership of the category) such as the fact that they generally do not have purple hair.

An obvious counter-argument to the thesis that prediction errors or violations are always negatively valenced is the existence of 'pleasant surprise', e. g. when one receives an offer that is better than expected. However, even the latter case, there is some evidence in humans and monkeys that the initial reaction to prediction error or surprise is generally negative (however short-lived) (Knight, Klepac, & Kralik, 2013; Noordewier & Breugelmans, 2013). We will later argue that the potential for a positive reaction to surprise is increased because the initial unexpectedness allows for a transition from high to lower prediction error (Huron, 2006; Van de Cruys

& Wagemans, 2011b). In brief, the agent-level emotion of surprise encompasses more than a single, momentary informational event (prediction error) at some level of the brain. As we will see, the dynamics of the failures and successes in prediction are more important here.

There are good biological reasons for why prediction confirmation should be good, while failures should be bad. A negative stimulus has only really deleterious consequences for survival if the system was not able to adequately prepare for it, by marshaling the necessary compensatory mechanisms (if necessary by acting to avoid it). Once this is taken care of, what could be a challenging or even threatening stimulus for bodily integrity, becomes pretty harmless. In fact, even a beneficial stimulus such as food could, for an unprepared body, be unpleasant. Conversely, predicting (and hence preparing to) a future negative stimulus that turns out not to occur, is often very wasteful for an organism. Hence, uncertainty or unpredictability is at the core of what an emotion is (Anselme, 2010; Bradford, Shapiro, & Curtin, 2013). The predictive machinery, while evolved to meet these adaptive challenges, is exploited to cover much more in humans. Here, pleasure for confirmation and displeasure for prediction violation can concern conceptual constructions that may not have life threatening (or facilitating) consequences, as a wealth of evidence in social psychology attests (reviewed in Heine, Proulx, & Vohs, 2006). This spawned a multitude of models going from cognitive dissonance theory to 'uncertainty management' model, the 'meaning maintenance' model or the 'system justification' model (Proulx et al., 2012).

In perception too, a positive appreciation for predictable stimuli is regularly reported, for example concerning familiar configurations (Ogawa & Watanabe, 2011), symmetric patterns (Evans et al., 2012; Makin, Pecchinenda, & Bertamini, 2012), or by correctly recognized objects (Chetverikov & Filippova, 2014; Jepma et al., 2012; Muth & Carbon, 2013; Winkielman & Cacioppo, 2001). Cognitive conflicts (prediction errors), on the other hand, have been shown to be aversive. A study by Dreisbach and Fischer (2012) showed that incongruent Stroop stimuli, as opposed to congruent ones, can prime people to more quickly evaluate negative words or pictures than positive ones (an indirect measure of negative affect; see also the replication by Schouppe et al., 2014). Importantly, the mere presentation of these conflict stimuli is sufficient, as participants were not instructed to actually perform the Stroop task. Neurally, perceptual prediction errors elicit activity in the habenula, a brain region found to be involved in processing negatively valenced ('worse than expected reward') prediction errors (Schiffer et al., 2012; Schiffer &

Schubotz, 2011). Other studies show that ('neutral') perceptual prediction errors also cause activity in striatal and midbrain regions, usually connected to reward/punishment and motivational functions (den Ouden et al., 2010; den Ouden, Friston, Daw, McIntosh, & Stephan, 2009; Iglesias et al., 2013). Together, these studies suggest that prediction error is never really neutral and that, as we will argue, emotion and motivation are intrinsically dependent on prediction errors.

But at this point one might object that positive emotion (or reward) as defined so far is too 'conservative' a concept[2]: we basically aim to return to familiar, overlearned states or situations and resist anything that deviates from those expected states (the mismatch or dissonance). This approach may explain the familiarity bias (the mere exposure effect) that is often reported (Lauwereyns, 2010), but it does not even remotely seem to capture our *experience* of reward in general. We easily get bored (loss of reward value) with very familiar or repetitive stimuli. More so, we seem to actively explore departures from well-trodden paths and expected situations. How do we explain that our motivations often lie outside of predictable ruts? And how can we more fully account for rewards as hedonic, pleasurable experiences derived from these different situations? That is what we will discuss next.

6.2.2 *Prediction error dynamics and affect*

Remember that we redefined drives as prediction errors (deviation from expected or desired states). This suggests a new way to look at rewards as well. Rewards are directly dependent on drives in the sense that the reward value of, say, a drop of water depends on the internal drive state of the organism (e. g., a thirsty rat). It is easy to see that what is critical then is the change in prediction errors (drive states). Rewards are given by a decrease of prediction errors, while punishment can be equated with an increase in prediction errors.

We propose that the affective valence is determined by the change in (or first derivative of) prediction error over time (Joffily & Coricelli, 2013; Van de Cruys & Wagemans, 2011b)[3], with positive valence linked to active reduction of prediction errors, and negative to increasing prediction errors. This makes sense because these temporal dynamics signal whether the organism is making progress (or

2 But it is far from a passive notion: to keep the organism within some expected range of a variable often means elaborate and vigorous activity of an organism (Egbert, 2013)

3 Note that the model by Joffily & Coricelli strictly speaking is not about prediction errors but rather about the more general concept of (variational) free energy

regress) in predicting its environment, which in the long term translates in proper functioning of the processes of life (fitness) (Damasio, 2003). The reward value of food evidently very much depends on how large the prediction error initially (i. e., how hungry) was, and hence how big a change the food consumption induced, but we propose this is a general pattern.

This goes beyond the simple view that prediction confirmation results in positive affect, while violations of predictions are negative. Once homeostasis, rather than being reactive, relies on predictive models, errors often do not have direct effect on homeostasis (or fitness). It becomes, then, equally important to monitor prediction error dynamics, as it is to monitor the errors as such. Mere presence of instantaneous prediction error does not seem to be an adequate basis of emotional valence. Positive affect might still occur for a large instantaneous error as long as this error is (or has been) in the process of being reduced.

Importantly, emotional valence is not something added to these error dynamics, it *is* those dynamics. They are a reflection of quality of processing, so they do not have to be evaluated in turn. As we saw earlier, humans are able to explicitly represent valence, but then this concept of valence obeys to predictive coding in the sense that it is itself a prediction explaining patterns of input. Explicit representation of valence may give these agents an edge in terms of their ability to adapt to changing environments, but it is not needed for an agent to have emotional valence. We connect positive and negative affect here to general purpose processing characteristics, detached from particular utility or motivations. They are purely determined by how the organism interacts with its environment (see also, Polani, 2009).

It is no stretch for humans to imagine that making progress in predicting various sensorimotor domains can be very rewarding. More challenging is to show those 'informational' rewards in non-human animals. However, Bromberg-Martin and Hikosaka (2011) have managed to show that monkeys too, are prepared to work to receive cues that reduce their uncertainty (reduce errors), even though their choice had no influence whatsoever on the actual reward subsequently received. The animals even chose the information cue more consistently than they typically choose a high probability reward over a low probability reward (Niv & Chan, 2011). Moreover, these informational gains elicited dopaminergic neural activity in midbrain regions similar to that for conventional rewards. Our account would predict that such effects generalize to other animal species, but of course, for there to be changes in prediction errors there need to be predictions formulated. Therefore, the spe-

cific instances of predictive gain will depend on the kind of models (predictions) an animal constructs about its world.

Even though the current view entails that emotions can arise wherever errors are compared, there are good computational and ecological reasons why change in errors is computed and compared within the limits of one and the same input domain. Comparing errors from very different perceptual levels or sensorimotor situations would be very demanding to the system, and, more importantly, unproductive. As Oudeyer et al. (2007) (p. 8) remark with regard to an artificial agent, such a system may "attribute a high reward to the transition between a situation in which a robot is trying to predict the movement of a leaf in the wind (very unpredictable) to a situation in which it just stares at a white wall trying to predict whether its color will change (very predictable)." Predictive coding proposes that specialization (functional segregation) in the brain stems from conditional independence of different representations —representations that have predictive relations organize into domains with tight interconnections (Friston, Schwartenbeck, et al., 2013; Stansbury, Naselaris, & Gallant, 2013). This architecture may also be used to evaluate changes in errors relating to predictions that are actually related.

The fact that these dynamics are subject to learning certainly makes behavioral testing of these ideas challenging. But although we cannot directly measure the predictions a participant's brain formulates for some presented perceptual input, we can either make sure people acquire new predictions for given stimuli, or use decreases in prediction errors that can be reasonably expected to be present and resolved by an experimenter's intervention. Suggestive evidence comes from a recent study looking at the affective consequences of conflict resolution (Schouppe et al., 2014). These authors build on the priming study by Dreisbach and Fischer (2012) discussed above, reporting that, while incongruent stimuli are aversive, once they are successfully solved more positive affect will follow than for congruent stimuli. The original prediction error (conflict) seems conducive to later reward from resolution, consistent with what we propose here.

Predictive progress has already been used to understand and implement intrinsic rewards in the domain of artificial intelligence (Oudeyer et al., 2007; Schmidhuber, 2010). More recently a decrease in prediction errors (or equivalently a predictive learning gain) was assumed to underlie intrinsic rewards in humans as well (F. Kaplan & Oudeyer, 2007). Agents that at each point try to maximize predictive progress, will avoid losing time in regions of sensorimotor space that are too difficult to predict with the current

capacities and regions that do not contain any learnable differences anymore, either because the domain is known or because what is left is noise variation. Hence, they will automatically focus on situations and stimuli that contain learnable differences, just above their current state of predictive knowledge, where the largest gain can be made. This guiding principle enables the agent to explore and proceed through stages of increasing predictive difficulty ('developmental phases'). There is some debate about the extent to which such an imperative to maximize prediction error reduction and PEM are one and the same thing (Clark, 2013b; Froese & Ikegami, 2013; Little & Sommer, 2013). Proponents of the 'maximizing learning gain' position contend that an organism driven by PEM will seek a dark room and stay there, because prediction error is maximally reduced there. However, as we elaborately discussed, a dark room is not actually a maximally expected situation (or does not stay so for long) in a generalized predictive coding framework. Prediction errors are always computed relative to an agent's possibly very complex, embodied model, with its specific organism-defining expectations, quickly rendering the dark room unexpected. While this seems to answer the 'negative' objection (why not stay in the dark room), can PEM also fully account why we humans 'positively' seek out prediction errors? This seems to depend on the kind of multilevel and second-order predictions we generate. As an example, if, at an abstract level, you expect yourself to be friendly, confirmation of this prediction will sometimes entail prediction errors on other, possibly lower levels. The key is to predict the violations as well, such that their impact can be reduced (see discussion on precision above). Similarly, if you expect to be a good darts player, you will need to tolerate some lower level sensorimotor errors to get there, usually because you can also (second order) reasonably expect the errors encountered to be reducible, based on previous experience.

6.2.3 *Non-conceptual metacognition*

Any operation performed on the prediction errors can be considered a form of metacognition. Similar to precision, the temporal comparison of prediction errors is a second-order operation. In the first-order process, prediction errors are information *in* the system (used to update predictions), while in the second-order process they become information *for* the system (Karmiloff-Smith, 1994), information that, we argue, is phenomenally experienced as valence and that may become available for processes beyond the predictive chain that created the errors. The result is a form of nonconceptual information about uncertainty that increases or decreases in the current

situation. It is not about the content but about the content-forming processes. The thesis here is that emotions are the qualitative experience (quale) of this kind of nonconceptual information. A related view is defended by Reisenzein (2009) in his belief-desire theory of emotion. He argues that emotions non-conceptually convey important changes in experiencer's belief system:

> "In representing these changes, emotional experiences inform us at the same time about the fate of our desires and beliefs while we acquire new knowledge about the world, and about the current state of the world-as-known in relation to our preexisting beliefs and desires." (Reisenzein, 2009, p. 221)

This nonconceptual information is available in terms of the affective tone. Note the connection with the concept of cognitive or perceptual fluency (Reber et al., 2004) as it pertains to liking and disliking of stimuli. It indicates the ease with which stimulus material is processed. Fluency should also be seen as a metarepresentation (Alter & Oppenheimer, 2009) and is arguably well characterized as the experience of actively reducing prediction errors (and disfluency as increases in prediction errors). If one identifies emotion with the *way* of processing rather than end-products (content of processing), perceptual (dis)pleasures and 'proper' emotions might be subsumed under the same principles. Specifically, fluency with regard to approaching high-level goals or biological concerns (bodily expected states) is what we usually associate with emotions. This idea is barely new. In a very influential control-theoretic approach to emotions, Carver and Scheier (1990) linked dynamics in mismatch between goals and perceptual inputs to dynamics in emotion. They described how multi-level goals should be interpreted as hierarchical reference values, from abstract idealized goals (e. g., having a self-image of a good person at the highest level), to more concrete actionable expectations (e. g., shoveling snow off of walks). In predictive coding terms, actions have to make sure that the agent can harvest the inputs that conform to 'trickled down' expectations. So, analogously to predictive coding, these expected values can generate errors at every level. Our own actions (or external circumstances) cause changes over time in discrepancies relative to these values. Carver and Scheier already argued that emotion is about monitoring the rate of discrepancy (prediction error) reduction, as we propose above. However, their analysis suggests a pertinent extension of what we presented so far. They suggest that the rate of mismatch reduction is in turn subject to a control loop, comparing actual with expected rate of change. Only when the current rate of prediction

error reductions deviates from the expected rate of reduction, so Carver and Scheier argue, one experiences emotion. This will of course be positive affect if the rate of progress to the goal is higher than expected, negative if it is lower than expected. Based on predictive coding, this makes a lot of sense. As we described, prediction error minimization is the way we perceive, so we are reducing errors all the time, e. g., when we successfully use our sensorimotor system to walk the street. Generally, little positive or negative emotion is involved despite these constant error reductions. This may mean that these sensorimotor changes in errors are not large enough, but most likely what rate is substantial depends on the expected rate of reduction for the current sensorimotor context. Where do the predictions of rates come from? These might very well be contextually learned through the same predictive machinery as for 'first-order' predictions, but a genetic contribution is not unthinkable. Individual differences in expected rates of error reduction, may account for certain dispositional affective traits. Indeed if the predicted rate of progress is set too high, an individual will tend to experience more negative affect than positive, because the prediction will rarely be matched (Carver & Scheier, 1990). This may happen, even if this person's actual rate of progress is very high. Furthermore, if the expected rates of progress are indeed at least partly learned specifically for different sensorimotor situations, this may constitute a form of emotion regulation. Specifically, the system may, through updating the expected rates, remain within a given range of emotional experience by adapting this criterion of expected rate of change (the neutral point).

Looking back, we have first encountered reasons to attribute emotion to prediction errors (mismatch) or confirmation as such, then we have shown it may be better attributed to changes over time in prediction errors, and finally to errors about expected rates of change. Importantly, these three can be independent. Borrowing an analogy from Carver and Scheier (1990); if we make make the parallel with distance, speed (first derivative of distance over time) and acceleration (second derivative), we can see that any rate of progress can be associated with any instantaneous prediction error, and further any change in rate of progress can co-occur with any instantaneous rate. The rate of error minimization seems to provide the necessary signal for valence. However, in mammals, especially humans, rate may be subjected to predictions of its own, moving important emotional dynamics to that level. Still, rate may determine the hedonic tone of what could be called 'background emotions' (Damasio, 2000). A steady rate of progress may induce a diffuse feeling of well-being, a sense of properly functioning bodily and

sensorimotor systems, akin to what is sometimes described as experience of *flow* (Csikszentmihalyi & Csikzentmihaly, 1991). Salient changes in emotions (i. e., emotions as commonly understood) on the other hand, seem linked to changes in rate of progress.

Emotions are notoriously volatile, comparative, and subject to habituation (Frijda, 1988). These characteristics are of course easily explained within the current framework. By definition, prediction errors and their temporal dynamics are dependent on learning. Pleasures from increased rates of predictive progress only last as long as this progress is possible. F. Kaplan and Oudeyer (2007) notes that 'progress niches are nonstationary'. Meanwhile, the contrastive property of emotion entails that a suboptimal state (sizable prediction error) may be pleasurable depending on the starting position, the rate of error reduction and the expected rate of reduction. Emotions emerge as perpetually moving targets in a system that grounds them in predictive dynamics, but note that this system does not aim to maximize the frequency of positive affect (nor would that be particularly adaptive) (Carver & Scheier, 1990). Rather, it may redistribute frequencies of positive and negative affect so as to preserve the range. Hence, one could loosely formulate a law of conservation of emotional range, from its negative to its positive pole (though individual differences in this range exist, see above).

Once an agent can track and learn to expect certain rates of change in prediction errors, it arguably will show a distinct propensity to explore and learn (follow very itinerant paths in state space). As such, this may form another counter-argument for the dark room objection against the principle of PEM. There will never be a stationary stimulus or situation satisfactory for an agent that expects some non-zero rate of prediction error minimization. We could even speculate that a high expected rate of error reduction may promote a stronger reliance on learning and culture.

6.3 AFFECT ELICITORS

One might object that the view we propose runs the risk of 'intellectualizing' emotions. Indeed we essentially described affect as a specific form of cognition (see also Duncan & Barrett, 2007). But this does not imply a cognitivistic theory of emotion. Error dynamics are common to all processing, be it interoceptive, exteroceptive, abstract goal-related or low-level sensorimotor. Errors are general processing products, so they include but go beyond those that concern fundamental homeostatic expected values. This may better account for the very broad range of situations that can engender positive or negative affects. For example, apart from biologically relevant

things, positive emotions may be experienced from scary movies, acquired tastes such as piquant foods (Rozin & Kennel, 1983) or painful stimulation such as masochistic pleasures (C. Klein, 2014). These instances may be difficult to explain from the viewpoint that pleasure is only attached to biologically instrumental situations (or appetitively conditioned stimuli). Below we review those emotions, from subtle to intense, that are usually considered to be atypical, for that reason. We try to show that, when taking into account the error dynamics relative to (learned) expected states, they are very representative emotions.

Development is a rich source of emotions. For example, the baby that wants to keep on playing peek-a-boo (Parrott & Gleitman, 1989) till predictions of object constancy are fully formed and the situation contains virtually no dynamics in prediction errors anymore. Or the child that is excited to hear the same bed-time story again and again, until errors are driven down by learning its structure. This is not just a matter of learning the plot until it is completely predictable, toddlers may enjoy learning (error dynamics) of lower level sensory regularities even more, as is for example clear from their preference for repetitive rhymes.

Usually these and related emotions in human or animal development are thought to be products of an evolved circuit for (emotion of) *play* (e. g., Panksepp, 2005). The adaptive function of such a system would be stimulating exploration of new, possibly more profitable options, and rehearsing sensorimotor and cognitive abilities that are vital in actual hazardous situations. Other emotion theorists emphasize the centrality of the underappreciated emotion of *interest*, for development and beyond (Izard, 2007; Silvia, 2001). The two factors that have been shown to determine interest can easily be translated to our approach, arguably gaining some specificity in the process. First, only new, unexpected or complex stimuli ('novelty-complexity' appraisal) can elicit interest (Silvia, 2008), implying that prediction errors are required [4]. The second factor is roughly described as comprehensibility (Silvia, 2008), an appraisal of one's capacity to deal with or understand the unexpected stimulus. In our terms, this would be an expectation of a positive rate of error reduction for the current sensorimotor context. It should be clear that, according to the predictive coding view, both appraisals can be made automatically and implicitly, meaning that interests can develop without much deliberate thought. In fact, we constantly make minute decisions on what to look at and what to engage in —our 'micro-interests'.

4 Note that complexity is also relative to the generated predictions

One of the most influential views on this topic is Berlyne's optimal level account of curiosity and exploration (Berlyne, 1970). He argued that organisms seek out stimuli with medium level complexity or novelty, to keep their arousal at an optimal, pleasing level. This preference for optimal level of complexity is corroborated in experiments with infants that looked longer at stimulus items that were neither very simple nor very complex (Kidd, Piantadosi, & Aslin, 2012). Rats too, prefer to spend time in arms of a maze of which the patterns on the walls were slightly more complex relative to the walls they preferred earlier (Dember, Earl, & Paradise, 1957). The latter studies emphasize the crucial role of experience, which can lower complexity (increase predictability). We would argue that organisms are very much tuned to reducible uncertainty in input. They explore stimuli with medium levels of prediction errors, because they predict a positive rate of error reduction in these inputs. Indeed, they have had experience of error reduction with slightly simpler but similar inputs. In agreement with this, 18 month old children already attend longer to learnable compared to unlearnable linguistic grammar, strongly suggesting they make good estimates of predictive (learning) progress they will be able to make (Gerken, Balcomb, & Minton, 2011). In adults, through experience, these dynamics, and the pleasures or displeasures derived from them, are not so much situated on the purely perceptual level, but rather on the conceptual level, e. g., stories, jokes or soaps. Although a complete treatment of social emotion will not be given here, observe that they often involve as convergence or divergence in opinions or 'worldviews' (expected states and beliefs). We make models of ourselves and others, like we do for the rest of our environment (Moutoussis et al., 2014), so similar error dynamics are at work in this context.

This view is somewhat related to previous theories about the 'need to resolve uncertainty', most aptly formulated by Kagan (1972) but going back to the writings of Festinger (1962) and Berlyne (1970). More recently, Biederman and Vessel (2006) described humans as 'infovores', to highlight that we derive pleasure from gathering information. However, there are several differences worth emphasizing. First, prediction errors are well-defined concepts, computed at different levels. Hence, this allows a more general formulation that spans from mere perceptual prediction errors to discrepancies of predictions on the level of (potentially conscious) goal and plans. For example it accommodates subtle perceptual emotions linked to discovery (or recovery) of organization or Gestalt where prediction errors are actively explained away by new interpretations that render the input more predictable (see above). On the behavioral front,

the rewarding sense of (em)power(ment) can be interpreted as the result of actively bringing about anticipated sensory effects through action execution (Herbart, 1891). Second, in contrast to previous related approaches, ours centers on prediction errors (rather than *any* uncertainty), more specifically their dynamics (rather than static uncertainty). Third, conceptualizing this drive as the 'wish to know' (Kagan, 1972) or the 'need for cognition' (Cacioppo & Petty, 1982) seems to suggest that this capacity is exclusive to so-called higher animals. Obviously, the more complex the predictive models an animal can learn and generate, the more prediction error dynamics matter. But as mentioned, we would object to an overly cognitivist interpretation of this 'drive', let alone one that is aimed at finding out some 'ground truth'. Knowledge captured in the models is always subjective, relative to predictions the system generates. It is also inherently constructivist (Heylighen & Joslyn, 2001) —the agent has no direct access to the 'real' world, but can only 'negotiate' its conditions by actively predicting (constructing) its characteristics.

Uncertainty and unpredictability have long been acknowledged as important factors in emotion and motivation (Anselme, 2010; Miller, 1981; Whalen, 2007). For example, rats seem more motivated to work for a reward in conditioning experiments that introduce some uncertainty in the predictive link between conditioned stimulus and unconditioned stimulus (reward) (Anselme, Robinson, & Berridge, 2013; Robinson, Anselme, Fischer, & Berridge, 2014). This seems consistent with our idea that rewards derive their rewarding capacity from reductions in (pre-existing) prediction errors. Beneficial effects of a sense of control ('mastery') (D. C. Klein & Seligman, 1976; Schulz, 1976; Turiano, Chapman, Agrigoroaei, Infurna, & Lachman, 2014) may similarly be explained as positive affect from a high expected rate of error reduction. Conversely, the exacerbating effect of uncontrollability and unpredictability on stress and anxiety is well-documented (Foa, Zinbarg, & Rothbaum, 1992; Hirsh et al., 2012; Jackson, Nelson, & Proudfit, 2014; Sarinopoulos et al., 2010). Evidently, this fits well with our emphasis on increasing prediction errors (and higher than expected rates of prediction error increases) as source of negative affect, although this hypothesis can be more directly tested by more systematically manipulating increases in unpredictability. There is some discussion about the extent to which (un)controllability effects can be explained by (un)predictability (Mineka & Hendersen, 1985), but in the current account the distinction largely dissolves —actions (to exercise control) are predictions as well, with concomitant expected levels of prediction error decrease or increase.

Typically, psychologists try to derive different particular rewards and punishments, including intrinsic ones, from (organism-specific) basic drives. However, it is hard to reduce intrinsic rewards or affects related to uncertainty and information to these primitives, as has already been remarked several decades ago (Kagan, 1972). The error dynamics approach makes the reverse movement: a single computational principle can be applied to organism-specific drives and to learned predictions (explaining uncertainty related emotion, including 'high-end' drives for knowledge). The need for coherent meaning ('epistemic coherence') and uncertainty reduction is, according to this approach, fundamental and emerges from biological principles. But emotions are attached to the dynamics of attaining this predictive coherence or error reduction.

In general, the positive emotional mark on unexpected progress towards predicted states is stronger (than 'just' progress). This is consistent with the view proposed here, that a higher than expected rate of error reduction determines positive emotion. This is best illustrated in laughter. In a poignant analysis, Sroufe and Waters (1976) observe that laughter results when a rapid, maximal tension build-up is followed by a rapid 'release' or 'recovery'. The ill-defined term 'tension' was often used to denote some incongruity in perceptual input, assumed to cause some negative arousal. Of course, prediction error can take its place, again gaining not only specificity, but also integration in a plausible theory of cognitive processing. Still more important to stress is that a steep, sudden gradient of prediction error will lead to a prediction of low rate of error reduction. If errors can in fact be reduced, e. g. through an appeal to different predictions (restructuring of input), the reduction rate will be much higher than expected, resulting in intensely positive affect (laughter). This is the typical processing profile, not only for peek-a-boo-like fun in children, but for instances of humor in general. For example, many jokes conform to an AAB pattern, with two similar events followed by a dissimilar one (Rozin, Rozin, Appel, & Wachtel, 2006). Evidently, one repetition is the minimum to set up a prediction in the listener. Once this is violated in the third event, this makes room for predictive progress that is unexpected. Consistent with our approach, both the gradients and the unexpectedness are crucial.

Emotions from music and arts comprise another area that is not readily explained from the point of view of traditional 'biologically-relevant' emotions. Rather, there is a very divided literature with one group of studies stating that aesthetic emotions derive from 'good Gestalts', symmetry, simplicity and predictability, while for the other group of studies these emotions stem from cognitive incon-

gruity, the breaking of patterns and expectation violations (Gaver & Mandler, 1987). We think this apparent contradiction can be resolved by realizing that the above described processes are at play —that strongly positive aesthetic emotions involve a transition from a state of higher to a state of lower prediction error (Van de Cruys & Wagemans, 2011b). Artists can use existing (culturally dominant) predictions or establish new predictions (e. g. rhythm or motif in music). Expectancy violations that they subsequently add, may allow for higher, indeed unexpectedly high, rate of error reduction. Another way to put this is that artists use errors to delay understanding, to allow their public to experience as if it is the first encounter with perceptual input, when the biggest progress could still be made.

This game of reinforcing, breaking and resolving predictions is more tractable in music, because phases are more clearly separable in time. However, in static, visual art similar processes seem to take place albeit largely 'offline', in the head of spectators. If the static snapshot of the art work is considered, it makes sense to talk about a kind of optimum level of perceptual or conceptual predictability in the work, but this ignores the processing dynamics taking place in the viewers. Also, note that artists mostly do not deliberately search for this sweet spot. They often push the limits in terms of unexpectedness, and then have to rely on a cultural learning process to bring unpredictability to a level that is optimal with regard to reducibility. Great artists are often not particularly popular in their own time.

The concept of fluency, that we equated with positive error reduction rate, has been invoked to explain aesthetic appreciation of visual art (Reber et al., 2004), but, according to our reasoning, *unexpected* fluency should be particularly potent (Topolinski, 2010). For example, visual art goes beyond the purely perceptual characteristics —often a 'violation' on the perceptual, stylistic level enables unexpected error minimization on the content level (for examples, see Van de Cruys & Wagemans, 2011b).

In a sense, even science is subject to these dynamics, for example in the often criticized craze for (allegedly) counter-intuitive findings in psychology, especially in top journals. These kinds of findings seduce, because they hold promise for great progress in understanding, namely by reducing prediction errors that were created by the findings. However, this also illustrates the fallibility of estimations of predictive progress, based on superficial characteristics. Predictive progress does not necessarily imply approaching the truth, it is always about progress relative to some maintained prediction (or set of interconnected predictions).

In conclusion, the motivational system we propose is more general than the accounts of play, interest, curiosity, humor or aesthetics described above, but can encompass all of them. It does not set these affects apart from other, allegedly more basic emotions (e. g., fear, happiness, etc.), because the error dynamics determine valence in both categories. Neither is more primary than the other. It should not surprise us then that neuroimaging data shows that more abstract stimuli (in interest, music, humor, etc.) still partly recruit the same neural areas as those involved in biologically relevant stimuli (e. g., Blood & Zatorre, 2001). Rather than that, as is usually argued, these stimuli somehow tap into the reward or emotion systems aimed at dealing with biologically relevant situations, the current view suggests they may all rest on the same mechanism. It explains a form of striving, an actively approaching to expected states, in all these activities, be it on a merely perceptual level, on a level of learned goals and expectations or on a bodily, interoceptive level. Impediments to this striving are apparent for the system as increasing prediction errors, and for the agent as negative affect. Moreover, the presence of these prediction errors will lower the expected future progress rate, potentially facilitating subsequent higher than expected progress rates. Attempts since Plato's Philebus, to interpret all pleasure as pleasure from relief from some unpleasantness or pain (from error), can also be seen in this light.

So far, we emphasized the commonality in emotions to the extent that they all have a valence component explained by error dynamics. We still need to examine what predictive coding can say about how emotions differentiate, for example into the various negative emotions such as shame, anxiety, etc. But before we do so, we briefly accentuate a few commonalities with the main currents in emotion theorizing.

6.4 EMOTIONS AND FEELINGS

Current theorizing on emotions is dominated by three accounts (each coming in various flavors). Several key ideas from these accounts can be mapped onto predictive coding concepts.

We can be relatively brief about a first category of emotion theories, often denoted as the 'basic emotion view'. The central idea here is that there is a limited set of evolutionarily defined fundamental emotions (such as fear, happiness, anger, etc.). Very scarce evidence for these 'natural kinds' of emotions exists at behavioral, physiological and neural levels. For example, most neural regions seem not uniquely linked to emotion, let alone specific 'basic' emotions (Barrett, 2006). More interestingly for our purposes, the brain

does not even honor the classical strict boundary between emotion and cognition. For example, where the amygdala was once argued to be a fear module (basic emotion) (Öhman & Mineka, 2001), it later became apparent that it is also involved in positive emotions (E. A. Murray, 2007), and even more broadly in cognitive and attentional tasks (Davis & J, 2001; Said, Dotsch, & Todorov, 2010). Conversely, signals reflecting emotional relevance and value have been found in primary sensory cortices (Serences, 2008; Stolarova, Keil, & Moratti, 2006). This is at least consistent with an account, such as the one proposed here, that does not separate emotions as a strictly different kind of processing completely detached (anatomically and functionally) from the rest of cognition.

A second main theory of emotions is the core affect theory. It can be traced back to Wundt (1897) and holds that all emotions have at their core only two fundamental ('primitive') dimensions, namely valence (pleasure-displeasure) and activation (arousal) (Barrett, 2006; Russell, 2003). This core affect compound is subsequently conceptualized (cognitively interpreted) to produce the multifarious feelings we experience. Core affect should be seen as an incessant "neurophysiological barometer of the individual's relationship to an environment at a given point in time." (Duncan & Barrett, 2007). Related to this conception, Frijda (2006, p. 82) notes: "pleasure is the positive outcome of constantly monitoring one's functioning". We can see a link here to prediction error dynamics, which are indeed a form of feedback on the system's own functioning as it deals with external and internal challenges. Proponents of the core affect theory have been somewhat vague about what exactly elicits or causes core affect. This is partly because core affect has no definite start, rather it is a continuous component of cognition. It is not something that is 'turned off' outside of the overt, intense emotional episodes to which emotions are often limited in everyday language. Core affect is a kind of continuous commentary on the system's processing of internal and external milieu. Consistent with this, prediction errors are assumed to be the general currency of processing in the brain, so they appear on the different hierarchical levels for different modalities and both for action and perception. Hence, temporal progress in errors can be evaluated in all these levels and domains. But even if affect is always present, it should be possible to indicate what elicits major *changes* in core affect (Lindquist, 2013). In our view, it is the change in rate of prediction errors reduction that elicit changes in the valence component of emotions.

A third and last category of emotion theories consists of appraisal theories (Arnold, 1960; Scherer, Schorr, & Johnstone, 2001). In its most basic form appraisal theories argue that emotions are formed

when a certain stimulus or situation is cognitively appraised using a series of 'checks'. Appraisals are the causal connection between stimuli and emotions, with similar appraisals causing similar emotions. Some important appraisals that have been distinguished are the appraisal of novelty (or unpredictability), pleasantness, goal relevance, and coping potential (Ellsworth & Scherer, 2003). Mark the appearance of a pleasantness appraisal. While anyone would agree this to be a central component of emotion, it cannot be used as an explanans for emotion, precisely because it is an explanandum here. One could say it is a hidden homunculus (an internal appraiser or evaluator) in this theory of emotion (see above), only hidden because such an operation seems unproblematic and self-evident for us, proficient appraisers. The other appraisal dimensions are more informative. That novelty or unpredictability appraisal is intrinsically linked to prediction errors needs no explanation. Goal relevance appraisals can, to the extent that goals can be cast as predictions (see above), also be construed as prediction errors, i. e., deviations from expected states. Finally, coping potential has a close connection with prediction errors and changes therein. The errors are a direct measure of the adequacy of our model of the world and the efficiency (cost) of processing. More specifically, we continually probe our coping potential by predicting performance (sensory consequences of actions) and computing errors. In fact, making progress (actively reducing errors) in predicting a certain activity domain would be a good indicator of adequate coping potential in this domain in the near future. Hence, the importance of expected rates of progress.

Do we gain anything from rephrasing appraisals as prediction errors? We think we do. Firstly, we can get rid of any lingering homunculi remnants. Secondly, this reframing naturally accounts for the multilevel character of appraisals (see discussion on multilevel predictions), while keeping the cognitive operation involved in appraisals simple enough to be executed quickly and automatically. The latter defuses an important criticism to appraisal theories, namely that the posited appraisal processes should be performed instantly and automatically (unconsciously) to figure as causal precedents of emotions (Moors, 2010). This seems far-fetched for some of the elaborate appraisal dimensions proposed in this literature[5]. In sum, the PEM view suggests that there may be an indefinite number of appraisals (as also suggested by Frijda, 2013), indeed as many as

5 Surely, more complex evaluations (e. g., of coping potential) are possible for us humans. However they are product of deliberate thought and hence not the most common causes of emotions.

there are predictions formed. But importantly, the error dynamics provide the crucial element for it to become an emotion.

6.4.1 *Constructing feelings*

Up to now, emotions as we defined them here are still far removed from the heterogeneous, object-directed, semantically and phenomenally rich, embodied feelings we usually experience. There is more to emotions than valence and intensity of valence, of course. However, to the extent that valence is a necessary ingredient of all emotions (we think it is, even for 'surprise'), the error dynamics seems to be at the core of all emotions. If we take for granted that prediction error minimization is all the brain does, what can we say about the rich distinctions between experienced feelings?

First, we have not said much about the embodied aspects of emotions, namely arousal and action tendencies. Both are obviously important components of emotion (Frijda, 1987), but rather than causally constitutive, we would put them at the output side. If, as we argued, emotions are caused by (unexpected) changes in prediction errors, these computations indeed seem especially important in tipping the balance from updating predictions —a strategy that may be inadequate when confronted with increasing, precise errors— to acting to change the things predicted. Arousal is then derived from such action preparations. Sure, dynamics in autonomic and action-related prediction errors can give rise to emotional valence as well, given that they are, as we discussed, governed by the same predictive coding principles. In fact, it seems that the closer to action or autonomic responses these error dynamics are situated, the more intensely negative or positive emotions induced by these dynamics are. This may, however, have more to do with the precision of the predictions than with discrete differences in weight or importance in these predictive systems. The expected flexibility in autonomic reference values is less than the flexibility of learned perceptual representations (predictions). However, only very rarely will stimuli directly violate expected autonomic values, rather they will violate predictions derived from these expected autonomic states. Moreover, just to avoid spending too much time in unexpected states, an organism will move away from expected autonomic states if it predicts that this will subsequently help a quicker return to expected state space regions. For this to work, all perception and action predictions have to be adequate and up-to-date. The counter-intuitive idea in this proposal is the lack of distinction between core concerns or needs of an organism versus just any predictions. But the main message we want to convey is that this difference is one in degree rather

than in kind. Also, note that in general much more uncertainty is attached to core concerns or high-level goals (e. g., they span more time than low-level perceptual predictions), so more possibility for unexpected gains and hence emotions.

Prediction of external stimuli is used in service of the body. It allows anticipation of what the body will need in terms of resources. Hence, it is important to accurately represent bodily states and their causes (Hohwy, 2011). However, just recruiting bodily resources, or representing bodily states (and their causes) isn't emotion. If the body perfectly predicts the need for resources based on external input and prior knowledge, there can be bodily activation (arousal) without much emotion. Again, momentary prediction errors do not imply much, it is the changes in (in this case somatovisceral) prediction error, especially the unexpected ones, that should lead to notable emotion. So, while we agree with the models by Seth (2013) and Gu, Hof, Friston, and Fan (2013) that hold that emotions (can) have to do with somatovisceral prediction errors, we stress that those accounts may not sufficiently explain the causes of emotions. The distinction should be clear: those accounts argue emotion is exactly like perception except of somatovisceral instead of exteroceptive inputs. Emotion is then inference to causes that explain (generate) somatovisceral inputs. We do not deny that these somatovisceral models are constructed, but focus on the dynamics in discrepancies of bodily state as causes of emotions. The origin of emotion lies not in being able to infer or predict (a cause of) bodily changes, but rather in how we succeed or fail to do so over time (error dynamics).

This brings us to the 'object-directedness' that characterizes and differentiates emotions. We inevitably attribute emotions to objects and events in the external world. It is, however, long known that this is a fallible process (e. g., Reisenzein, 1983; Schachter & Singer, 1962), in other words, misattribution is very common which may completely change the course of an emotion. The prototypical example is when negative arousal (e. g., of fear) can become positive if attributed to a positive event (e. g., an infatuation) (Dutton & Aron, 1974). Our account does not situate the origin of emotion in arousal, but it entails a different, more fundamental misattribution. Precisely the interpretation that emotions are entirely caused by objects out there is misguided, because in fact they are linked to processing characteristics (see also Reber et al., 2004) rather than content of processing itself. The evaluation of error dynamics seems to provide a parallel (affective) dimension to experience, that is not strictly linked to the content (predictions) or particular prediction errors taking part in those dynamics. Still, the specific and diverse

forms emotions can take, seem largely dependent on the conceptual context (sensorimotor or cognitive domain) in which the error dynamics appear. But such attributions are always constructions, they will never be directly about what caused the emotions, the (changes in) error reduction rates.

We can draw on the conceptual act theory of emotion for a very related solution to this problem of diversity of emotions, within a theory that only has few 'primitive' emotional ingredients. It holds that emotions are conceptualizations or interpretations of interoceptive and exteroceptive states (Barrett, 2014; Lindquist, 2013; Russell, 2003). Emotions are the result of an act of making meaning of a bodily state in the current conceptual context. Emotions do not have essences but are predictable constellations of interoceptive and exteroceptive stimuli that a child, during development, learns to label as a discrete emotion. The different discrete emotions are 'just' the result of categorizing one's state (interoceptive and exteroceptive). Barrett (2014) explains:

> "To say that it is an 'act' does not mean it is deliberate, but rather that it is not a passive event because the perceiver is not merely detecting and experiencing what it is out there in the world or what is going on inside his or her body —prior experiences (i. e., knowledge) play a role in directing sensory sampling, and in making meaning of the sensory array to create momentary experiences, whether or not they are creating an instance of emotion, perception, or cognition. The word 'act' merely signals that every mental event is, to a considerable extent, infused with top-down (perceiver-based) input." (Barrett, 2014, pp. 12-13)

Such a view is of course precisely what predictive coding attempts to formalize, specifically when applied to the domain of interoception (Seth, 2013). We would like to make a distinction here between feelings and emotions. Feelings are indeed a form of making sense of underlying emotions. This could already be seen as a form of coping, a reaction to emotion: categorizing or labeling an emotion to make it predictable (reducing errors). The underlying emotion consists, in our reasoning, of the deviation (errors) from expected error reduction rates. The first-order errors that determine these dynamics can be multisensory (combined interoceptive and exteroceptive). The feeling state will then be finding predictions that best explains the co-occurrence (regularity) of situational context (exteroceptive input) and bodily states, together with changes in rates of error reduction (second-order). Hence, the intentional content of feelings is the product of inferences, but the generation of

emotion lies in error dynamics. Differently put, if emotions are categorized, a kind of understanding is attained, which explains away (part of) the unexpected changes in errors, hence removing (some of) the emotionality (see also Wilson and Gilbert (2008) on 'affective adaptation').

To give one simple example of how a similar emotion might result in different feelings: an unexpected increase of prediction errors may be associated with both fear and shame (seeing that both are negative emotions). But on the basis of the different conceptual, situational context (e. g., shame probably concerns internalized social expectations, fear not necessarily so) they are differently interpreted and experienced. Interestingly, just by conceptualizing it as shame, we might activate coping strategies that in related situations helped returning to more expected states. The shame prediction for this constellation of inputs is predictive of certain actions or thought strategies that are in turn predictive of a reduction of prediction errors (e. g., actions to restore one's reputation with others).

Humans seem able to take the predictive machinery offline, simulating (counterfactual) predictions and running them against simulated data. How this happens exactly is still largely unexplored (but see Clark, 2013b; Hohwy, 2013). Given that powerful emotions can spring from imagining and thinking about events (Carver & Scheier, 1990; Russell, 2003), our account postulates that error dynamics similarly feature here. It even seems to allow us to experience a kind of vicarious progress (regress), for our social partners. This needs to be clarified in future work. In any case, the fact that changes in errors can be monitored in different sensorimotor, conceptual and even counter-factual domains could be key to understanding different forms of mixed and conflicting affects, or volatile emotions based on currently dominant predictive activity in the brain. Tentatively, this may go some way towards explaining the complexity of emotions.

6.4.2 *Function of affect*

We have seen how emotionality is always to some extent part of cognitive processing. It is a form of non-conceptual information that nonetheless has clear phenomenological quality. So, for example, positive affect is what it is like for the organism to make predictive progress. We will not address the thorny question whether the *conscious* positive or negative experience is in any way instrumental or just an epiphenomenon (Schultz, 2007). Could it be that pleasure is just a by-product of successful PEM? Once a function receives a circumscribed computational form, it is hard to see why it could not be

realized just as well without any phenomenal qualities. This is not an issue specific to our account, nor to PEM in general. Setting aside this 'hard problem', we should briefly discuss the possible functions of the kind of computations that, according to our approach, underlie affect.

These computations should be understood as building a model about how uncertainties evolve in the current context. It seems plausible that these models could be crucial in guiding choices (implicitly or explicitly) about whether to continue to engage with the current sensorimotor activities or whether to disengage and switch. Specifically, (unexpected) decreases in prediction errors should raise predictive engagement, in line with how emotion motivates us to remain involved in activities. Confidence, as a rather emotional form of (meta)cognition (Chetverikov & Filippova, 2014), also seems to stem from these processes. In contrast, (unexpected) increases of prediction errors should induce disengagement from current predictive activity in order to move to a more predictable 'set'. As we proposed earlier, it may also change the balance in favor of action to controlling input (conform to predictions; assimilation) instead of a continued search for revised predictions (accommodation). An action to avoid perceptual input (by averting the eyes), or even a mental switch to leave a certain way of thinking can also be ways to (temporarily) return to a more expected rate of error reduction. A more speculative idea is that the negative affect of increasing rates of prediction errors, motivates one to seek compensatory progress and order. Preliminary evidence for this idea has come from studies finding increased predictable pattern perception when confronted with ambiguity or lack of control (van Harreveld, Rutjens, Schneider, Nohlen, & Keskinis, 2014; Whitson & Galinsky, 2008). Presenting people with conceptual inconsistencies ('meaning threats') has also been found to increase motivation to find predictable structure in artificial grammar, and better learning of that grammar (Proulx & Heine, 2009). The domain of uncertainty (prediction errors) was always different from that of the newly found order, so it seems justified to call this 'compensatory'. Greenaway, Louis, and Hornsey (2013) found that the loss of control leads to more belief in paranormal precognition, and suggest that the loss of predictive frame is compensated by a belief in a predictable future. More broadly, uncertainty or inconsistencies may lead people to reaffirm their own (predictable) worldviews, such as nationality, ideology or religion (Inzlicht, Tullett, & Good, 2011; Proulx et al., 2010; Valdesolo & Graham, 2014). The negative affect thought to drive these effects is, according to our theorizing, a direct reflection of the higher than expected increase in prediction errors. These examples may then all

boil down to efforts to return to an expected, positive rate of uncertainty reduction.

Another function of these affective computations relates to learning and attention. Joffily and Coricelli (2013) formally show that the first derivative over time of prediction errors can fulfill a similar function as the one usually assigned to the precision mechanism (see above). An increase in prediction errors (negative valence) may indicate that actual, important changes in the world have taken place, so input (incoming prediction errors) should be weighted more heavily compared to top-down predictions (that apparently need to be updated). In other words, the error rates can be used as a meta-learning signal, tuning the learning rate for new inputs, depending on whether there is much too learn (i. e., in a changing world) or not. What we defined as expectations of error reduction rates then take on the role of expectations on learnability of particular input domains. These are models about what we do not know yet about the structure of the world (Joffily & Coricelli, 2013) and how these uncertainties will evolve, i. e., to what extent we estimate these uncertainties will be reducible. Joffily and Coricelli (2013) argue that a model that uses rate of change in errors is more parsimonious than one including precisions (a conventional PEM model), but more work will be needed to clarify both differences in computational realizability and biological plausibility.

Note that we are concerned here with sub-personal models of uncertainty. The fact that we maintain these stands in stark contrast to our difficulty in deliberate, explicit reasoning about uncertainties. In general, uncertainties in processing inputs seem shielded from our awareness. Our account suggests that they may still have some phenomenal characteristics, namely in the form of emotional valence.

This casts a different light on attention as well. It has always been difficult to separate behavioral relevance or value from attention in itself (Maunsell, 2004). Attention is about getting the reliable information, given the 'questions' (i. e., predictions) you have (Gottlieb, 2012). That is why expected precisions of prediction errors are assumed to underlie attention: expected precision can adapt the gain of input such that we attend to precise information about our current predictions. If instead of precision we use value as we defined it here, we always attend to valuable input in the sense of stimuli with high expected rate of error reduction. Such a treatment may explain why reward and attention seem to have unified selection mechanisms in primary visual cortex (Stanisor, van der Togt, Pennartz, & Roelfsema, 2013).

In short, our view may permit a reinterpretation of the amply documented effects of emotion on cognition, including perception,

attention (Vuilleumier, 2005), learning and social cognition. Such influences will need to be systematically explored in light of our hypothesis. To give but one example, positive mood seems to induce a greater reliance on prior or top-down knowledge, as indicated by an increased influence of prior judgments, scripts or stereotypes in event or person perception (Bless, 2000; Bodenhausen, Kramer, & Süsser, 1994). This dovetails with the proposed view that positive mood is linked to high predictive progress, implying that the models the organism has about its world have improved and so are adequate. A rational conclusion for the system would then be to increase reliance on these prior, top-down models (and reduce the influence of prediction errors).

6.5 NEURAL IMPLICATIONS

So far, we have no direct neural evidence for the existence of the proposed computational operations (or their products). In part, this is probably due to the difficulty of the question. There is only very preliminary and indirect evidence to date for the separable error and prediction populations of neurons (de Gardelle et al., 2012), as postulated by PEM. However, the general idea that there are different levels in the hierarchy, with separable prediction errors has recently received support (Diuk et al., 2013; Wacongne et al., 2011). Once we succeed in properly localizing those, on different hierarchical levels of processing, we can start looking for dynamics in these errors and neural populations or regions that track these changes and generate predictions of error reduction rates. It is likely that these computations are performed distributed in the brain (similar to first-order predictive coding), given the widely distributed encoding of uncertainty in the brain depending on the domain concerned (D. R. Bach & Dolan, 2012). But since we are talking about basic emotional value, which is presumably highly conserved throughout phylogenetic history, and shown to engage various subcortical regions, we need to look to those regions (including others thought to be involved in emotion processing) as well. In this regard, the overlap in regions found to be important in processing uncertainty and those active for emotional processing, is promising. Again, a full exploration is beyond the scope of this chapter, but if we take the insula for example, we find it to be engaged both in processing of (bodily aspects of) feelings and more broadly in uncertainty and risk processing (Preuschoff et al., 2008; Singer, Critchley, & Preuschoff, 2009). In an interesting case study, Picard (2013) describes two patients with seizures originating in insular regions, who experienced an exhilarating sense of bliss coinciding with a

strange sense of certainty. One of them reported that what he felt was "not unlike a continuous series of profound 'a-ha!' moments" (Picard, 2013, p. 2496). Picard suggests this may be due to the insula's role in processing prediction errors and uncertainty. We hypothesize more specifically that it may make temporal comparisons of prediction errors (or contains the products of those operations) such that it can represent predictive progress, and possibly expectations of such progress. One can then imagine that abnormal activity in this region can give rise to a generalized feeling that the world is infused with meaningfulness (value), because it is as if it perfectly reflects the agent's own beliefs about it, with complete confidence that anything beyond immediate reach can be resolved.

The anterior cingulate cortex, too, is activated by behavioral conflicts, inconsistencies, performance errors, and negative affect (e. g., see Shackman et al., 2011), but may also track reliability (reducibility) of prediction errors (Behrens et al., 2007). Whether those regions may actually estimate expected error increases or decreases, and whether this may explain their role in affective cognition, remains to be seen. It is a plausible hypothesis, especially given these regions' involvement in cognitive control (Mushtaq, Bland, & Schaefer, 2011), regulating (dis)engagement, switching, and attention as we discussed in the previous section. Crucially, emotion in general has a broad, diffuse impact on subsequent processing, and subcortical nuclei are well placed and connected to 'broadcast' signals of predictive progress or failure beyond their region of origin. Neuromodulators, such as norepinephrine, that seem to intervene in inference, as well as in emotion (Dayan, 2012b), may have an important role here.

Finally, Verduzco-Flores and O'Reilly (2014) present a model of the cerebellum functioning for motor control that computes increase of error (error derivative), rather than instantaneous error, to adjust reaching movements. Indeed, our account would predict that cerebellum, like the rest of the brain, would represent these changes, specifically relative to predictions about sensory consequences of one's actions, that are assumed to be encoded there (Miall, Weir, Wolpert, & Stein, 1993; Synofzik, Lindner, & Thier, 2008). Apart from possible functional roles in learning or attention (such as that described by Verduzco-Flores & O'Reilly, 2014), we contend that such error increases and decreases contribute to emotional valence. Incidentally, cerebellum damage leads to a particular cognitive-affective syndrome, characterized by executive and sensorimotor deficits as well as depressive symptoms (lack of emotion) and/or anxiety (Wolf, Rapoport, & Schweizer, 2009). We are reminded here that our approach may be able to shed new light on other psy-

chiatric disorders that combine cognitive and emotional problems, such as depression, autism, and schizophrenia. This seems a fruitful avenue for future research, considering the recently emerging predictive coding accounts of these mental illnesses (Adams et al., 2013; Gradin et al., 2011; Van de Cruys et al., 2014). For example, could depression be well-characterized as a generalized low expectation of error reduction rates (see also Huys & Dayan, 2009)?

6.6 CONCLUSIONS

Visual perception —the 'noblest sense'— has a prominent place in Western culture. Throughout our history, it was considered our main route to find the ground truth about the world out there. Particularly in the history of science, scholars relied on this allegedly direct connection to the material world, yet still keeping the proper objective distance. It became the eye that objectifies and masters (Jay, 1994), key to the Western world's most important technological breakthroughs. Note the *voir* ('to see') in the French *pouvoir* ('to be able to'; 'power') and *savoir* ('to know'). With the hegemony of the eye came the myth of the all seeing eye (Koenderink, 2014). The idea, still pervading modern science, that there is a (quasi-godly) vantage point from which one could see the objective, veridical states of affairs —those that science will uncover. This is, as Koenderink (2014) forcefully argues, of course a fata morgana. As we saw, physical differences in the world only become information (meaning) by the way we probe them, with our (organism-specific) predictions. This means that value and information are intertwined by construction —courtesy of our existence as biological organisms.

We have tried to present an account of emotional value starting from the premise that all the brain is doing can be summarized by generalized predictive coding and its PEM principle. Emotional valence appears here as a particular informational process, but not mere- ly Shannon-informational given the source of predictions and the particular engagement they represent. It shows some promise for the predictive coding framework to become a common, well-specified language for psychology. Although our view builds on several core ideas developed by prominent emotion theorists, reinterpreting emotions in PEM terms led to notions that depart from concepts originating in folk-psychology and sometimes carried through in emotion theory. This should be welcomed, because it leads us to question implicit assumptions in those concepts and it stresses that familiar, intuitively clear concepts (such as reward) do not necessarily correspond to unitary psychological and neural substrates.

This is only a first exploration of how emotions and related dimensions or forms of experience can be reconceptualized within predictive coding. We highlighted various aspects that deserve more attention. Many challenges lie ahead, but the question of whether the brain indeed tracks error increases and decreases and forms predictions about those, is open to empirical and computational investigation. Theoretically, there is a worry that PEM, as broadly conceived as we have done here, will not sufficiently specify the particular mechanisms of cognitive processing. The issue is even more pressing with regard to emotions, often conceived of as ad hoc, adaptive 'hacks' in the brain. Still, as we tried to make clear, every 'trick in the bag' is the result of predictive matching to environmental regularities. When such a quick-and-dirty, evolutionarily prepared strategy should be applied, rather than the general purpose, flexible —but costly— predictive system, is an arbitration that may also be solvable in a Bayesian way (based on relative uncertainties, Clark, 2013a). In reality, there will probably be a whole continuum of predictive flexibility, as we saw.

We really do need structure, so we can see we are nowhere.

— J. Cage

7

Conclusions

A few common strands run through all chapters of this work. One of them is that predictive coding is not nearly as new as one might conclude reading recent articles about it. Important precursors of the central ideas of predictive coding can be found in philosophy (e. g., James, 1890; von Glasersfeld, 1991), cybernetics (e. g., Bateson, 1979), biology (Rosen, 2012[1985]), and psychology (e. g., Gregory, 1980; Helmholtz, 1878; Köhler, 1920). This is a good thing because it adds to the plausibility of the framework. It does not mean that predictive coding is just old wine in new bags. Due to the influence from artificial intelligence and computational neuroscience, it has greater computational specificity (e. g., the precision mechanism) than its precursors. The fact that it is more committed to a particular architecture, makes it more amenable to empirical tests and simulations. Hence, a second common strand is that we could come up with new behavioral and modeling questions based on the account. A third is that prediction errors are fruitful building blocks to think about the pathological mind and about emotion. If errors are our way of 'meeting the world', they are indeed bound to determine our experience and our difficulties in interacting with this world. A fourth and last common strand is that predictive coding as a form of anticipatory processing allows us to fully appreciate the continuity of psychology with the rest of biology. This enabled us to start to respond to the fundamental question of how something like value could appear in a world of mere matter and information.

Three things remain for us here: taking stock of what we gained, looking ahead at what might be in store for the PEM framework, and briefly considering the recurrent concern about falsifiability of the framework. We will start with the latter.

The worry about falsifiability is of course not specific to our proposal, nor to predictive coding at large. For example, the larger family of Bayesian approaches has been accused of being unfalsifiable (Bowers & Davis, 2012). However, the argument seems mostly a

red herring. It is true that to make a broad framework such as predictive coding falsifiable one always requires auxiliary assumptions and initial conditions (Chalmers, 1999). This is no different for the theory of natural selection or the second law of thermodynamics, to give but a few examples of in principle unfalsifiable theories. Griffiths, Chater, Norris, and Pouget (2012) make the useful distinction between frameworks and models. Frameworks should be judged based on how productive they are in generating specific models and new empirical predictions (again, the pragmatic criterion rules), not on how falsifiable they are. The concrete models derived from it, however, are disprovable. They will be strengthened by empirical success and weakened if they need ad hoc constructions to explain data.

In the past chapters, we mainly used predictive coding as a general framework but it should be apparent that specific testable computational models derived from it can be and have been formulated (e. g., Iglesias et al., 2013; Mathys et al., 2011; Rao & Ballard, 1999). We now need to investigate whether they can be extended to cover the broad scope of tasks and faculties that predictive coding claims to explain. In any way, the opportunity is unique in the sense that psychologists, neuroscientists and computational modelers are all examining different aspects of the same framework. We already see this happening for predictive coding accounts of schizophrenia (Adams et al., 2013; Schmack et al., 2013; Stephan, Friston, & Frith, 2009) and we hope we have helped start such movement for ASD.

For both ASD and affect we have seen that, although the concepts are often compatible with existing theories, the advantage is that they are more integrated and specific in cognitive, computational and neural terms. Hence the ideas should in principle be more falsifiable than earlier accounts. However, we will not know until we have actually examined all its implications and assumptions. This large enterprise has indeed been started in the literature and the current work should be able to contribute to that discussion. Importantly, the neural implementation can be tested — multiple research groups (Bastos et al., 2012; Egner & Summerfield, 2013; Muckli, Petro, & Smith, 2013) set out the lines along which to do so. Similarly, the computational realizability can (and will need to) be evaluated (Blokpoel, Kwisthout, & van Rooij, 2012).

Most importantly, there is a healthy discussion about which computational scheme and information-theoretical objective the brain actually follows (e. g., Clark, 2013b; Phillips, 2013; Spratling, 2010; Thornton, 2014a), given that it performs some form of probabilistic, predictive processing. These proposals of guiding principles include regular PEM and maximization of predictive progress (Chapter 6),

but also the principle of maximizing mutual information (proposed by Little & Sommer, 2013), of maximizing the amount correctly predicted (proposed by Phillips, 2012) or the infotropic principle (proposed by Thornton, 2014a). Although all of these principles seem to explicitly or implicitly reduce prediction errors, the actual, full consequences of their application for perception and action can be hard to predict. These are not just theoretical debates; they will inform us about the need for explicit representation of prediction errors in separate units (Phillips, 2013; Rasmussen & Eliasmith, 2013) and about the possible architectures (e. g., do feedforward pathways only signal prediction error?) that can be used to realize predictive processing in the brain.

With that, we arrive at what we gained and what may still be in store for the presented views.

CHAPTER 2 Concerning our HIPPEA proposal on autism (Chapter 2), we tried to show that problems emerge throughout the behavioral literature in ASD when actual changes in input structure are mixed with accidental variability (noise), presumably because adaptive estimation of precision of input is missing. This is a more subtle deficit than previous theories assume, but accounts better for the mixed evidence of behavioral deficits in ASD. We have indicated several new (multidisciplinary) empirical questions that need no repeating. It *is* worth emphasizing that one of the most crucial issues is to try to make a clearer, empirically informative distinction between more precise prediction errors and less precise (weaker) predictions (or priors). Although they may have similar effects in some areas (e. g., a more heavy reliance on bottom-up evidence), they seem at least in principle dissociable when considering the result of inference (higher posterior precision for high precision prediction errors) and with regard to (meta)learning. This discussion may superficially seem just a rehash of the controversy between weak central coherence theory and enhanced perceptual functioning theory, but we gained a computational framework (predictive coding/Bayesian inference) that promises better specification and quantification of clinical alterations in cognitive tasks.

The approach provided fresh behavioral hypotheses, specifically with regard to (lack of) context-dependent tuning of precision of prediction errors. Sensorimotor or perceptual tasks manipulating context uncertainty in ASD are already on the way and should prove very informative for the validity of the approach. We cannot resist giving two more empirical avenues. A first one would be to use a visual search task, one of the clearest examples of enhanced performance in ASD. However, Ma, Navalpakkam, Beck, Berg, and Pouget (2011) showed an interesting variation of this task, using varying

reliabilities on a trial by trial basis. According to their findings, humans perform near optimal taking into account these uncertainties. HIPPEA would predict this would not hold for people with ASD, even though they might be better in basic visual search. A second research design concerns the problem in meta-learning —learning where learnable differences are situated. Gerken et al. (2011) found that TD infants of 17 months attend longer to learnable compared to unlearnable linguistic patterns. HIPPEA predicts that infants with ASD would not show this difference (or show it later for given materials), suggesting that this is an interesting candidate for an early behavioral marker.

As we described, the theory still leaves plenty of room for phenotypical variability. One might, for example, imagine that what is sometimes called 'core autism' is characterized by high and inflexible precision, while other individuals on the spectrum may only lack the context-dependent flexibility or, alternatively, only have abnormally high mean precision. These are clear hypotheses that lend themselves to empirical refutation. On the neural front, it remains to be clarified what the underpinnings of precision computations in the brain may be (we identified candidates) and whether these can eventually be correlated with symptom severity in ASD. Although we think the account fares well in comparisons with more established single deficit accounts, further discussions should shed light on whether it can comply to the criteria for such accounts as identified throughout the literature (e.g. universal, specific, mechanistic, consistent with neurobiological data, etc.; e. g., see Pellicano, 2011). Overall, we think our cognitive account can help connect neurophysiology of ASD with the clinical, experiential level of description.

An interesting discussion will surely continue on Bayesian models for ASD and other mental disorders. Will these approaches have the flexibility to adequately account for the different kinds of disordered mind, while also limiting the amount of variation that can be expected? We hope we have managed to highlight some pertinent questions that have to be clarified theoretically to arrive at a specific predictive coding or Bayesian account of ASD (and to make the distinction with schizophrenia). There is a sense in which any behavioral profile can be modeled with Bayes if one only (ad lib) picks the right priors, so a central question becomes whether there is inefficient model updating ('broken inference') or 'just' an abnormal model (prior beliefs) (Moutoussis et al., 2014). This distinction is further clouded by the fact that inefficient model updating will also lead to deviant priors. Generally, one could say that the Bayesian approach has too many degrees of freedom here. But as Griffiths et al. (2012) point out, this is just because these are more transparent

(hence salient) in this approach compared to other cognitive models. Indeed, the Bayesian account forces the user to make these assumptions explicit. This brings up a final important challenge for the field: bringing about close collaborations between researchers with strong technical know-how in model building, and those skilled in working with clinical populations on cognitive and behavioral tasks.

CHAPTER 5 Our study with the point-light walkers still left at least two possible interpretations open: people may differ in the prior or prediction they apply based on the behavioral relevance of the different priors. Or, they might differ in terms of the sensitivity with which they process bottom-up (in our case perspective) cues. Given what we saw about precision in Chapter 2, we could rephrase the latter as heightened precision of incoming evidence for people with high social anxiety. This may be a temporary, stimulus or context-related, higher precision, rather than a higher precision across the board as we suggested for ASD. There may be an important link here given that (social) anxiety is also commonly reported in ASD, as we discussed in Chapter 2. This idea is also compatible with reports of better change detection and more bottom-up driven attention for people with high trait social anxiety, as we already indicated in Chapter 5 (Moriya & Sugiura, 2012; Moriya & Tanno, 2009). Interesting in this regard is that stress potentiates early (EEG) stages of visual processing but attenuates later stages (Shackman et al., 2011). It is not far-fetched to assume that people with high trait social anxiety experienced more state anxiety in the social setting of the experiment (e. g., because it requires interacting with the experimenter in a the performance context), which may for them generally cause a higher precision of bottom-up input. We deem such general, stimulus-independent changes in input processing likely, so future studies should in any case include a neutral, geometrical (e. g., a structure-from-motion) bistable figure, and ideally an additional, lower level visual task. This would allow us to check whether not the content (value) of the figures but rather the way (any) presented input is processed differs between people with anxiety and without. Additionally, it would be important to verify whether state anxiety induction (in non-trait-anxious people) has the same effects as trait anxiety in these tasks. Inducing ecologically valid anxiety in the lab (without resorting to shocks) is not trivial but a game context with different threat levels seems promising (Mobbs et al., 2007).

Since we published our study, one related study (Heenan & Troje, 2014) appeared, also looking at the link between anxiety and the facing bias in bistable biological motion stimuli. These authors showed that reducing anxiety (by ten minutes of physical exercise or by mus-

cle relaxation) reduced the facing bias for biological walkers. They propose that reducing anxiety reduces the threat value of the facing percept, hence rendering the percepts more equal. These results seem inconsistent with ours, though note that they reduced (state) anxiety rather than making use of high trait social anxiety as we did. In their second experiment they did explicitly try to increase anxiety (using an autobiographical anxiety induction method), but this did *not* result in higher facing bias, as would be expected following the socio-emotional relevance hypothesis. Importantly, their manipulation worked fine (post-questionnaire showed increased anxiety), it just did not cause the expected increase in facing bias. In fact, if anything it seems to slightly reduce (non-significantly) it (see their figure 4). Also puzzling is that the manipulation checks (heart rate measurements and questionnaire before and after relaxation, induction or control) showed no reduction in anxiety in the relaxation condition (where they did find an effect in their perceptual task) but an increase in anxiety in the anxiety induction condition (where they did not find an effect in the perceptual task). Although this might be due to limitations of the measurements used, we should remain open to the possibility that their effects were not caused by differences in anxiety per se (this holds for our study as well). Maybe a cleaner test of the affective relevance hypothesis —the idea that the value of the prior biases perception— would be to use a neutral bistable structure-from-motion stimulus, but aversively condition one of the two percepts (counterbalanced). A bias towards or away from the conditioned interpretation would speak to the relevance hypothesis (a positivity or negativity bias), rather than one in terms of sensitivity to inputs or another domain-general process.

CHAPTER 3, 4 AND 6 The three chapters left can be treated together because they all explore our hypothesis on the relation between emotion dynamics and error dynamics. These ideas and particularly the accompanying empirical test are very preliminary. But much of the theoretical developments is compatible with existing theories on psychoaesthetics and emotion. With regard to the latter, our reasoning brought non-mainstream emotion accounts to the fore that, based on the logic of generalized predictive coding, deserve such prominent place. We foresee exciting work here, about how varying predictability influences (subtle) emotion dynamics. However, at least two obstacles may impede progress. The first is the noted individualized, experience-dependent nature of predictability. The second is that we are not (in the first place) concerned with agent-level prediction errors (surprise or expectation violation), but with sub-personal processing products. It means that we will even-

tually need to rely on neural measures for tracking these dynamics. Hence, the added value of our account largely depends on the success of predictive coding proponents in pinning down the physiological basis of the computational scheme.

There is another reason for optimism about these views. In a recent symposium on probabilistic inference in the brain, philosopher and predictive coding proponent Andy Clark made a plea for the rise of the PEM robots. Indeed, we could envision artificial prediction error minimizing agents, traversing their unique landscape of prediction error, creating new valleys and descending new slopes. One could test their ability to perceive, learn and execute adaptive behavior in toy worlds, initially of low complexity. In the current context it would additionally be interesting to see how animate and intrinsically motivated these agents seem. What if we add the capacity to form expectations about error reduction rates? We know from Heider and Simmel's (1944) work that people readily ascribe emotions and intentions to simple, moving stimuli, but the movements of these PEM robots would not be preprogrammed, but depend on internal dynamics. The agents would start with a set of expected states given some essential (homeostatic) variables and they should develop the ability to generate useful habits matching the structure of the environment. Some progress in this direction has already been made by Egbert and Canamero (2014). They created simple agents of the kind first proposed by Braitenberg (1986, hence called Braitenberg vehicles) that learned predictable sensorimotor sequences. These 'habits' could support staying within the viable range of essential variables and hence are self-maintaining. The specific form of these habits are, however, not predefined but developed autonomously and varied from agent to agent. Thornton (2014b), on the other hand, implemented a very simple form of 'seeking' behavior in these Braitenberg vehicles, using a predictive processing scheme. Note that both of these examples used a form of predictive processing, but not predictive coding per se. It will be interesting to see if one could build PEM vehicles, that incorporate homeostatic predictions but attain some independence in the kinds of habits they develop. Such robotics experiments could also teach us to what extent the different principles (see above) that could guide predictive processing are equivalent or have practical, implementational (dis)advantages. More generally, creating PEM-based agents could probably point us to shortcomings of the framework that we might not notice otherwise (Thornton, 2014b). By doing this, we would close the circle by applying the constructivist inspiration of predictive coding to the theory of predictive coding itself: only by

constructing PEM agents we can learn whether they could be viable models of living, perceiving, behaving and learning organisms.

If an emotion theory along the lines presented in Chapter 6 is right, we might be getting some formal grasp on affective value and intrinsic motivation, key characteristics of proactive, living organisms. The convergence between computational neuroscience and psychology as seen through this predictive coding account is encouraging. Although computational neuroscience largely leads the way here, we should not overlook the role of experimental psychologists and systematic students of phenomenology as indispensable sounding boards. Hopefully, the current work can serve as proof that predictive coding provides a fruitful way to think about the mind-brain, from low-level perceptual issues to emotional and existential issues. Because ultimately, PEM is about gathering evidence for one's own existence (Hohwy, 2014). This is the necessary cognitive counterpart of an organism's capacity to regenerate itself, through metabolism, as a unit distinguished from its surroundings (autopoiesis; Varela, Maturana, & Uribe, 1974).

I want to end this work on a word of humility. As scientists, we are used to deal in uncertainties. But we want our piece of the predictive progress goodness as much as the next fellow, if we think it is within reach. It can be hard not to be seduced by an orderly, unified (predictable!) theory of the mind. However, the only way of knowing whether our intuition is a rightful sense of predictive progress, is to test the limits of the theory, and explore its implications to the fullest. We have done part of this work with regard to affect and autism. If not true, one may hope that it will nonetheless help sharpen our questions and specify the explananda. The most we can hope for as scientists, is not to leave the field more cluttered for those who come after us —anything extra is a bonus. It may give intellectual pleasure to bring it all together in a few orderly principles, but we have to remind ourselves that nature, and especially the mind, is more stubborn and messy than we might like or can imagine.

Coda

If one were to devise some guidelines for well-being, based on the principles of (dis)pleasure we put forward in Chapter 6, these could go along the following lines. Find a cognitive niche where one can enjoy the trajectory (error reduction) rather than the fleeting results. This has its perils as well, as when people flee in video games, because the rate of progress made there is often incomparable to

daily life (because it is tailored to the current proficiency of the player). Another strategy could be to attempt to silence the predictions; because no errors without predictions. It is highly questionable whether this is at all possible, given that any mental activity is, according to PEM, predictive. Still, predictive activity may be focused to lower level perceptual ones, that can be confirmed more easily. In fact, this is part of what Buddhism and related meditative practices try to accomplish: focus on simple changes in interoceptive and exteroceptive inputs, without activating higher level conceptual, or even goal-related (cf. 'nonjudgmental') predictions about them. One could look at repetitive rituals of other religions in a similar way. They are not ways to avoid predictive activity (which would be impossible), but rather efforts to move predictive activity to more 'manageable' areas. This aligns with what we described in Chapter 6 with respect to religion as compensatory predictability. The human 'curse of cognition' (Vaneechoutte, 1993) may actually be a curse of prediction. Our explosive predictive capacity, pinnacling in full-blown 'foresight', bore an inability to remain in the present.

We can find some light however, in the current 'egalitarian' theory of pleasure. Pleasures come in various forms, but there are no 'higher' ones and 'lower' ones according to this view. Consistent with age-old wisdom, someone with objectively few (classical) rewards, could still have an extremely satisfying (rewarding) life. Conversely, someone with objectively lots of 'rewards' could have a dreadful life. Finally, insights that are very bleak, but increase our predictive grasp, can be deviously pleasurable.

Bibliography

Adams, R. A., Stephan, K. E., Brown, H. R., & Friston, K. J. (2013). The computational anatomy of psychosis. *Frontiers in Schizophrenia*, *4*, 47. doi: 10.3389/fpsyt.2013.00047

Albright, T. D. (2012). On the perception of probable things: Neural substrates of associative memory, imagery, and perception. *Neuron*, *74*(2), 227–245. doi: 10.1016/j.neuron.2012.04.001

Alderson-Day, B., & McGonigle-Chalmers, M. (2011). Is it a bird? is it a plane? Category use in problem-solving in children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *41*(5), 555–565. doi: 10.1007/s10803-010-1077-9

Alink, A., Schwiedrzik, C. M., Kohler, A., Singer, W., & Muckli, L. (2010). Stimulus predictability reduces responses in primary visual cortex. *The Journal of Neuroscience*, *30*(8), 2960 –2966. doi: 10.1523/JNEUROSCI.3730-10.2010

Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: Role of the STS region. *Trends in Cognitive Sciences*, *4*, 267–278. doi: 10.1016/S1364-6613(00)01501-1

Alpers, G. W., & Gerdes, A. B. M. (2007). Here is looking at you: emotional faces predominate in binocular rivalry. *Emotion*, *7*(3), 495–506. doi: 10.1037/1528-3542.7.3.495

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review*, *13*(3), 219–235. doi: 10.1177/1088868309341564

Amaral, D. G., Schumann, C. M., & Nordahl, C. W. (2008). Neuroanatomy of autism. *Trends in Neurosciences*, *31*(3), 137–145. doi: 10.1016/j.tins.2007.12.005

Amaral, J. L., Collins, S., Bohache, K. T., & Kloos, H. (2012). Beyond the Black-and-White of autism: How cognitive performance varies with context. In H. Kloos (Ed.), *Current topics in children's learning and cognition.* InTech.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5.* Arlington, Va.: American Psychiatric Association.

Anderson, C. J., & Colombo, J. (2009). Larger tonic pupil size in young children with autism spectrum disorder. *Developmental Psychobiology*, *51*(2), 207–211. doi: 10.1002/dev.20352

Anselme, P. (2010). The uncertainty processing theory of motivation.

Behavioural Brain Research, *208*(2), 291–310. doi: 10.1016/j.bbr.2009.12.020

Anselme, P., Robinson, M. J. F., & Berridge, K. C. (2013). Reward uncertainty enhances incentive salience attribution as sign-tracking. *Behavioural Brain Research*, *238*, 53–61. doi: 10.1016/j.bbr.2012.10.006

Apps, M. A. J., & Tsakiris, M. (2014). The free-energy self: A predictive coding account of self-recognition. *Neuroscience & Biobehavioral Reviews*, *41*, 85–97. doi: 10.1016/j.neubiorev.2013.01.029

Aragon. (1971). *Henri Matisse: roman*. Gallimard.

Arnal, L. H., & Giraud, A. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences*, *16*(7), 390–398. doi: 10.1016/j.tics.2012.05.003

Arnheim, R. (1974). *Art and visual perception*. Berkeley, CA: University of California Press.

Arnold, M. B. (1960). Emotion and personality.

Ashwin, E., Ashwin, C., Rhydderch, D., Howells, J., & Baron-Cohen, S. (2009). Eagle-eyed visual acuity: an experimental investigation of enhanced perception in autism. *Biological Psychiatry*, *65*(1873-2402 (Electronic)), 17–21. doi: 10.1016/j.biopsych.2008.06.012

Asperger, H. (1944). Die "autistischen psychopathen" im kindesalter. *Archiv fur Psychiatrie und Nervenkrakheiten*, *117*, 76–136.

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus Coeruleus-Norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, *28*(1), 403–450. doi: 10.1146/annurev.neuro.28.061604.135709

Augustin, M. D., Carbon, C., & Wagemans, J. (2012). Artful terms: A study on aesthetic word usage for visual art versus film and music. *i-Perception*, *3*(5), 319–337. doi: 10.1068/i0511aap

Aviezer, H., Trope, Y., & Todorov, A. (2012). Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science*, *338*(6111), 1225–1229. doi: 10.1126/science.1224313

Bach, D. R., & Dolan, R. J. (2012). Knowing how much you don't know: a neural organization of uncertainty estimates. *Nature Reviews Neuroscience*, *13*(8), 572–586. doi: 10.1038/nrn3289

Bach, D. R., Hulme, O., Penny, W. D., & Dolan, R. J. (2011). The known unknowns: Neural representation of Second-Order uncertainty, and ambiguity. *Journal of Neuroscience*, *31*(13), 4811–4820. doi: 10.1523/JNEUROSCI.1452-10.2011

Bach, M., & Dakin, S. C. (2009). Regarding "Eagle-Eyed visual acuity: An experimental investigation of enhanced perception

in autism". *Biological Psychiatry*, *66*(10), e19–e20. doi: 10.1016/j.biopsych.2009.02.035

Baird, G., Simonoff, E., Pickles, A., Chandler, S., Loucas, T., Meldrum, D., & Charman, T. (2006). Prevalence of disorders of the autism spectrum in a population cohort of children in south thames: the special needs and autism project (SNAP). *Lancet*, *368*(9531), 210–215. doi: 10.1016/S0140-6736(06)69041-7

Bakouie, F., Zendehrouh, S., & Gharibzadeh, S. (2009). Does a kind of over-fitting occur in the brain of autistic patients? *The Journal of neuropsychiatry and clinical neurosciences*, *21*(3), 343. doi: 10.1176/appi.neuropsych.21.3.343

Balcetis, E., & Dunning, D. (2010). Wishful seeing. *Psychological Science*, *21*(1), 147–152. doi: 10.1177/0956797609356283

Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, *11*(7), 280–9. doi: 10.1016/j.tics.2007.05.005

Bar, M. (2009). The proactive brain: memory for predictions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1235–1243. doi: 10.1098/rstb.2008.0310

Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2007). Threat-related attentional bias in anxious and nonanxious individuals: A meta-analytic study. *Psychological Bulletin*, *133*(1), 1–24. doi: 10.1037/0033-2909.133.1.1

Baron-Cohen, S. (2001). Theory of mind and autism: A review. *International Review of Research in Mental Retardation, Vol 23*, *23*, 169–184.

Baron-Cohen, S. (2006). The hyper-systemizing, assortative mating theory of autism. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *30*(5), 865–872. doi: 10.1016/j.pnpbp.2006.01.010

Baron-Cohen, S., Ashwin, E., Ashwin, C., Tavassoli, T., & Chakrabarti, B. (2009). Talent in autism: hyper-systemizing, hyper-attention to detail and sensory hypersensitivity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1471-2970 (Electronic)), 1377–1383. doi: 10.1098/rstb.2008.0337

Baron-Cohen, S., Richler, J., Bisarya, D., Gurunathan, N., & Wheelwright, S. (2003). The systemizing quotient: an investigation of adults with asperger syndrome or high–functioning autism, and normal sex differences. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*(1430), 361–374. doi: 10.1098/rstb.2002.1206

Baron-Cohen, S., Ring, H., Bullmore, E., Wheelwright, S., Ashwin,

C., & Williams, S. (2000). The amygdala theory of autism. *Neuroscience & Biobehavioral Reviews*, *24*(3), 355–364. doi: 10.1016/S0149-7634(00)00011-7

Barrett, L. F. (2006). Are emotions natural kinds? *Perspectives on Psychological Science*, *1*(1), 28 –58. doi: 10.1111/j.1745-6916.2006.00003.x

Barrett, L. F. (2014). The conceptual act theory: A précis. *Emotion Review*, *6*(4), 292–297. doi: 10.1177/1754073914534479

Barrett, L. F., & Bar, M. (2009). See it with feeling: affective predictions during object perception. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1521), 1325–1334. doi: 10.1098/rstb.2008.0312

Barrett, L. F., Mesquita, B., & Gendron, M. (2011). Context in emotion perception. *Current Directions in Psychological Science*, *20*(5), 286–290. doi: 10.1177/0963721411422522

Bartz, J. A., Zaki, J., Bolger, N., & Ochsner, K. N. (2011). Social effects of oxytocin in humans: context and person matter. *Trends in Cognitive Sciences*, *15*(7), 301–309. doi: 10.1016/j.tics.2011.05.002

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron*, *76*(4), 695–711. doi: 10.1016/j.neuron.2012.10.038

Bates, D., Mächler, M., & Bolker, B. (2012). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software (forthcoming)*.

Bateson, G. (1979). *Mind and nature: A necessary unity*. Dutton New York.

Bauman, M. L., & Kemper, T. L. (1994). Neuroanatomic observations of the brain in autism. *The neurobiology of autism*, 119–145.

Bauman, M. L., & Kemper, T. L. (2005). *The neurobiology of autism*. JHU Press.

Baxter, M. G., & Murray, E. A. (2002). The amygdala and reward. *Nature Reviews Neuroscience*, *3*(7), 563–573. doi: 10.1038/nrn875

Bechtel, W. (2009). Explanation: Mechanism, modularity, and situated cognition. *The Cambridge handbook of situated cognition*, 155–170.

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*(9), 1214–1221. doi: 10.1038/nn1954

Behrmann, M., Peterson, M. A., Moscovitch, M., & Suzuki, S. (2006). Independent representation of parts and the relations between them: Evidence from integrative agnosia. *Journal of Experimen-*

tal Psychology: Human Perception and Performance, *32*(5), 1169–1184.

Bellon, M. L., Ogletree, B. T., & Harn, W. E. (2000). Repeated storybook reading as a language intervention for children with autism: A case study on the application of scaffolding. *Focus on Autism and Other Developmental Disabilities*, *15*(1), 52–58. doi: 10.1177/108835760001500107

Belmonte, M. K., Cook, E. H., Anderson, G. M., Rubenstein, J. L. R., Greenough, W. T., Beckel-Mitchener, A., ... Tierney, E. (2004). Autism as a disorder of neural information processing: directions for research and targets for therapy*. *Molecular Psychiatry*, *9*(7), 646–663. doi: 10.1038/sj.mp.4001499

Ben-Sasson, A., Hen, L., Fluss, R., Cermak, S., Engel-Yeger, B., & Gal, E. (2009). A meta-analysis of sensory modulation symptoms in individuals with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *39*(0162-3257 (Print)), 1–11. doi: 10.1007/s10803-008-0593-3

Berlyne, D. E. (1966). Curiosity and exploration. *Science*, *153*(3731), 25–33. doi: 10.1126/science.153.3731.25

Berlyne, D. E. (1970). Novelty, complexity, and hedonic value. *Perception & Psychophysics*, *8*(5), 279–286. doi: 10.3758/BF03212593

Bertone, A., Mottron, L., Jelenic, P., & Faubert, J. (2003). Motion perception in autism: a "complex" issue. *Journal of Cognitive Neuroscience*, *15*(0898-929X (Print)), 218–225. doi: 10.1162/089892903321208150

Bertone, A., Mottron, L., Jelenic, P., & Faubert, J. (2005). Enhanced and diminished visuo-spatial information processing in autism depends on stimulus complexity. *Brain*, *128*(10), 2430–2441. doi: 10.1093/brain/awh561

Bialek, W., Nemenman, I., & Tishby, N. (2001). Predictability, complexity, and learning. *Neural Computation*, *13*(11), 2409–2463. doi: 10.1162/089976601753195969

Biederman, I., & Vessel, E. (2006). Perceptual pleasure and the brain. *American Scientist*, *94*(3), 247. doi: 10.1511/2006.3.247

Bird, G., Leighton, J., Press, C., & Heyes, C. (2007). Intact automatic imitation of human and robot actions in autism spectrum disorders. *Proceedings of the Royal Society B-Biological Sciences*, *274*(1628), 3027–3031. doi: 10.1098/rspb.2007.1019

Bird, G., Silani, G., Brindley, R., White, S., Frith, U., & Singer, T. (2010). Empathic brain responses in insula are modulated by levels of alexithymia but not autism. *Brain*, *133*(5), 1515–1525. doi: 10.1093/brain/awq060

Bishop, D. V. M. (1989). Autism, asperger's syndrome and semantic-

pragmatic disorder: Where are the boundaries? *International Journal of Language & Communication Disorders*, *24*(2), 107–121. doi: 10.3109/13682828909011951

Bitbol, M. (2007). Ontology, matter and emergence. *Phenomenology and the Cognitive Sciences*, *6*(3), 293–307. doi: 10.1007/s11097-006-9041-z

Blackford, J. U., Allen, A. H., Cowan, R. L., & Avery, S. N. (2012). Amygdala and hippocampus fail to habituate to faces in individuals with an inhibited temperament. *Social Cognitive and Affective Neuroscience*. doi: 10.1093/scan/nsr078

Blake, R., & Shiffrar, M. (2007). Perception of human motion. *Annual Review of Psychology*, *58*(1), 47–73. doi: 10.1146/annurev.psych.57.102904.190152

Blakemore, S., Tavassoli, T., Calo, S., Thomas, R., Catmur, C., Frith, U., & Haggard, P. (2006). Tactile sensitivity in asperger syndrome. *Brain & Cognition*, *61*(0278-2626 (Print)), 5–13. doi: 10.1016/j.bandc.2005.12.013

Bless, H. (2000). The interplay of affect and cognition: The mediating role of general knowledge structures. In *Feeling and thinking: The role of affect in social cognition* (pp. 201–222). New York, NY, US: Cambridge University Press.

Blokpoel, M., Kwisthout, J., & van Rooij, I. (2012). When can predictive brains be truly Bayesian? *Frontiers in Psychology*, *3*. doi: 10.3389/fpsyg.2012.00406

Blood, A. J., & Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proceedings of the National Academy of Sciences*, *98*, 11818–11823.

Bodenhausen, G. V., Kramer, G. P., & Süsser, K. (1994). Happiness and stereotypic thinking in social judgment. *Journal of Personality and Social Psychology*, *66*(4), 621–632. doi: 10.1037/0022-3514.66.4.621

Bögels, S. M., & Mansell, W. (2004). Attention processes in the maintenance and treatment of social phobia: hypervigilance, avoidance and self-focused attention. *Clinical Psychology Review*, *24*(7), 827–856. doi: 10.1016/j.cpr.2004.06.005

Boll, S., Gamer, M., Gluth, S., Finsterbusch, J., & Büchel, C. (2013). Separate amygdala subregions signal surprise and predictiveness during associative fear learning in humans. *The European Journal of Neuroscience*, *37*(5), 758–767. doi: 10.1111/ejn.12094

Bölte, S., Holtmann, M., Poustka, F., Scheurich, A., & Schmidt, L. (2007). Gestalt perception and Local-Global processing in High-Functioning autism. *Journal of Autism and Developmental Disorders*, *37*(8), 1493–1504. doi: 10.1007/s10803-006-0231-x

Bölte, S., Hubl, D., Dierks, T., Holtmann, M., & Poustka, F. (2008). An fMRI-study of locally oriented perception in autism: Altered early visual processing of the block design test. *Journal of Neural Transmission*, *115*(3), 545–552. doi: 10.1007/s00702-007-0850-1

Bölte, S., Schlitt, S., Gapp, V., Hainz, D., Schirman, S., Poustka, F., ... Walter, H. (2012). A close eye on the eagle-eyed visual acuity hypothesis of autism. *Journal of Autism and Developmental Disorders*, *42*(5), 726–733. doi: 10.1007/s10803-011-1300-3

Bonnel, A., McAdams, S., Smith, B., Berthiaume, C., Bertone, A., Ciocca, V., ... Mottron, L. (2010). Enhanced pure-tone pitch discrimination among persons with autism but not asperger syndrome. *Neuropsychologia*, *48*(1873-3514 (Electronic)), 2465–2475. doi: 10.1016/j.neuropsychologia.2010.04.020

Bonnel, A., Mottron, L., Peretz, I., Trudel, M., Gallun, E., & Bonnel, A. (2003). Enhanced pitch sensitivity in individuals with autism: a signal detection analysis. *Journal of Cognitive Neuroscience*, *15*(0898-929X (Print)), 226–235. doi: 10.1162/089892903321208169

Booth, R., Charlton, R., Hughes, C., & Happe, F. (2003). Disentangling weak coherence and executive dysfunction: planning drawing in autism and attention-deficit/hyperactivity disorder. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *358*(0962-8436 (Print)), 387–392. doi: 10.1098/rstb.2002.1204

Bornstein, R. F. (1989). Exposure and affect: Overview and meta-analysis of research, 1968–1987. *Psychological Bulletin*, *106*(2), 265–289. doi: 10.1037/0033-2909.106.2.265

Bornstein, R. F., Kale, A. R., & Cornell, K. R. (1990). Boredom as a limiting condition on the mere exposure effect. *Journal of Personality and Social Psychology*, *58*(5), 791–800. doi: 10.1037/0022-3514.58.5.791

Botvinick, M., & Cohen, J. (1998). Rubber hands 'feel' touch that eyes see. *Nature*, *391*(6669), 756. doi: 10.1038/35784

Bourdieu, P. (1984). *Distinction: a social critique of the judgement of taste*. Harvard University Press.

Bowers, J. S., & Davis, C. J. (2012). Bayesian just-so stories in psychology and neuroscience. *Psychological Bulletin*, *138*(3), 389–414. doi: 10.1037/a0026450

Bradford, D. E., Shapiro, B. L., & Curtin, J. J. (2013). How bad could it be? alcohol dampens stress responses to threat of uncertain intensity. *Psychological Science*, *24*(12), 2541–2549. doi: 10.1177/0956797613499923

Braitenberg, V. (1986). *Vehicles: Experiments in synthetic psychology*.

MIT Press.

Brincker, M., & Torres, E. B. (2013). Noise from the periphery in autism. *Frontiers in Integrative Neuroscience*, *7*. doi: 10.3389/fnint.2013.00034

Brock, J. (2012). Alternative Bayesian accounts of autistic perception: comment on Pellicano and Burr. *Trends in Cognitive Sciences*, *16*(12), 573–574; author reply:574–575. doi: 10.1016/j.tics.2012.10.005

Bromberg-Martin, E. S., & Hikosaka, O. (2011). Lateral habenula neurons signal errors in the prediction of reward information. *Nature Neuroscience*, *14*(9), 1209–1216. doi: 10.1038/nn.2902

Brooks, A., Schouten, B., Troje, N. F., Verfaillie, K., Blanke, O., & van der Zwan, R. (2008). Correlated changes in perceptions of the gender and orientation of ambiguous biological motion figures. *Current Biology*, *18*(17), 728–729. doi: 10.1016/j.cub.2008.06.054

Brown, C., Gruber, T., Boucher, J., Rippon, G., & Brock, J. (2005). Gamma abnormalities during perception of illusory figures in autism. *Cortex*, *41*(3), 364–376. doi: 10.1016/S0010-9452(08)70273-9

Brown, J., Aczel, B., Jiménez, L., Kaufman, S. B., & Grant, K. P. (2010). Intact implicit learning in autism spectrum conditions. *Quarterly Journal of Experimental Psychology*, *63*(9), 1789–1812. doi: 10.1080/17470210903536910

Brown, S., Gao, X., Tisdelle, L., Eickhoff, S. B., & Liotti, M. (2011). Naturalizing aesthetics: Brain areas for aesthetic appraisal across sensory modalities. *NeuroImage*, *58*(1), 250–258. doi: 10.1016/j.neuroimage.2011.06.012

Burack, J. A. (1994). Selective attention deficits in persons with autism: Preliminary evidence of an inefficient attentional lens. *Journal of Abnormal Psychology*, *103*(3), 535–543. doi: 10.1037/0021-843X.103.3.535

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*(1), 116–131. doi: 10.1037/0022-3514.42.1.116

Carroll, L. S., & Owen, M. J. (2009). Genetic overlap between autism, schizophrenia and bipolar disorder. *Genome Medicine*, *1*(10), 1–7. doi: 10.1186/gm102

Carruthers, P. (2008). Meta-cognition in animals: A skeptical look. *Mind & Language*, *23*(1), 58–89. doi: 10.1111/j.1468-0017.2007.00329.x

Carver, C. S., & Scheier, M. F. (1990). Origins and functions of positive and negative affect: A control-process view. *Psychological Review*, *97*(1), 19–35. doi: 10.1037/0033-295X.97.1.19

Cerra, P. L., & Bingham, R. (1998). The adaptive nature of the human neurocognitive architecture: An alternative model. *Proceedings of the National Academy of Sciences*, *95*(19), 11290 –11294. doi: 10.1073/pnas.95.19.11290

Chalmers, A. F. (1999). *What is this thing called science?* (3edition ed.). Indianapolis: Hackett Publishing Co.

Chater, N. (1996). Reconciling simplicity and likelihood principles in perceptual organization. *Psychological Review*, *103*(3), 566–81. doi: 8759047

Chater, N., & Vlaev, I. (2011). The instability of value. In M. Delgado, E. A. Phelps, & T. W. Robbins (Eds.), *Decision making: Attention and performance XXIII* (p. 81–100).

Chetverikov, A., & Filippova, M. (2014). How to tell a wife from a hat: Affective feedback in perceptual categorization. *Acta Psychologica*, *151*, 206–213. doi: 10.1016/j.actpsy.2014.06.012

Cheung, C., Yu, K., Fung, G., Leung, M., Wong, C., Li, Q., ... McAlonan, G. (2010). Autistic disorders and schizophrenia: Related or remote? an anatomical likelihood estimation. *PLoS ONE*, *5*(8), e12233. doi: 10.1371/journal.pone.0012233

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., & Schultz, R. T. (2012). The social motivation theory of autism. *Trends in Cognitive Sciences*, *16*(4), 231–239. doi: 10.1016/j.tics.2012.02.007

Chopin, A., & Mamassian, P. (2010). Task usefulness affects perception of rivalrous images. *Psychological Science*, *21*(12), 1886 –1893. doi: 10.1177/0956797610389190

Chopin, A., & Mamassian, P. (2011). Usefulness influences visual appearance in motion transparency depth rivalry. *Journal of Vision*, *11*(7). doi: 10.1167/11.7.18

Clark, A. (2011). Predictive coding. *Edge World Question Center 2011*.

Clark, A. (2013a). The many faces of precision. *Frontiers in Theoretical and Philosophical Psychology*, *4:*, 270. doi: 10.3389/fpsyg.2013.00270

Clark, A. (2013b). Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(03), 181–204. doi: 10.1017/S0140525X12000477

Conant, R. C., & Ashby, R. W. (1970). Every good regulator of a system must be a model of that system. *International Journal of Systems Science*, *1*(2), 89–97. doi: 10.1080/00207727008920220

Connors, S. L., Crowell, D. E., Eberhart, C. G., Copeland, J., Newschaffer, C. J., Spence, S. J., & Zimmerman, A. W. (2005). beta2-Adrenergic receptor activation and genetic polymorphisms in autism: Data from dizygotic twins. *Journal of Child Neurology*, *20*(11), 876–884. doi: 10.1177/

08830738050200110401

Constantino, J. N. (2002). *The social responsiveness scale*. Los Angelos: Western Psychological Services.

Cook, R., Brewer, R., Shah, P., & Bird, G. (2013). Alexithymia, not autism, predicts poor recognition of emotional facial expressions. *Psychological Science*. doi: 10.1177/0956797612463582

Cooney, R. E., Atlas, L. Y., Joormann, J., Eugène, F., & Gotlib, I. H. (2006). Amygdala activation in the processing of neutral faces in social anxiety disorder: Is neutral really neutral? *Psychiatry Research: Neuroimaging*, *148*(1), 55–59. doi: 10.1016/j.pscychresns.2006.05.003

Courville, A. C., Daw, N. D., & Touretzky, D. S. (2006). Bayesian theories of conditioning in a changing world. *Trends in Cognitive Sciences*, *10*(7), 294–300. doi: 10.1016/j.tics.2006.05.004

Craig, A. D. B. (2009). How do you feel - now? the anterior insula and human awareness. *Nature Reviews Neuroscience*, *10*(1), 59–70. doi: 10.1038/nrn2555

Crespi, B. (2013). Developmental heterochrony and the evolution of autistic perception, cognition and behavior. *BMC Medicine*, *11*(1741-7015 (Electronic)), 119. doi: 10.1186/1741-7015-11-119

Csikszentmihalyi, M., & Csikzentmihaly, M. (1991). *Flow: The psychology of optimal experience* (Vol. 41). HarperPerennial New York.

Dakin, S., & Frith, U. (2005). Vagaries of visual perception in autism. *Neuron*, *48*, 497–507.

Damasio, A. (2000). *The feeling of what happens: Body and emotion in the making of consciousness*. Harvest Books.

Damasio, A. (2003). *Looking for spinoza. joy, sorrow, and the feeling brain*. Orlando, Austin, New York, San Diego, Toronto, London: Harcourt, Inc.

Darwin, C. (2011). *Metaphysics, materialism, and the evolution of mind: The early writings of charles darwin* (Subsequent edition ed.). Chicago: University Of Chicago Press.

Davis, M., & J, P. (2001). The amygdala: Vigilance and emotion. *Molecular Psychiatry*, *6*(1), 13–34. doi: 10.1038/sj.mp.4000812

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, *8*(12), 1704–1711. doi: 10.1038/nn1560

Dawson, G. (2008). Early behavioral intervention, brain plasticity, and the prevention of autism spectrum disorder. *Developmental Psychopathology*, *20*(1469-2198 (Electronic)), 775–803. doi: 10.1017/S0954579408000370

Dawson, G., Finley, C., Phillips, S., Galpert, L., & Lewy, A. (1988).

Reduced p3 amplitude of the event-related brain potential: its relationship to language ability in autism. *Journal of Autism and Developmental Disorders*, *18*(0162-3257 (Print)), 493–504.

Dawson, M., Mottron, L., & Gernsbacher, M. A. (2008). Learning in autism. In *J. H. Byrne (Series Ed.) & H. Roediger (Vol. Ed.). Learning and memory: A comprehensive reference: Cognitive psychology of memory* (pp. 759–772). London: Elsevier.

Dayan, P. (2012a). How to set the switches on this thing. *Current Opinion in Neurobiology*, *22*(6), 1068–1074. doi: 10.1016/j.conb.2012.05.011

Dayan, P. (2012b). Twenty-five lessons from computational neuromodulation. *Neuron*, *76*(1), 240–256. doi: 10.1016/j.neuron.2012.09.027

Dayan, P., Kakade, S., & Montague, P. R. (2000). Learning and selective attention. *Nature Neuroscience*, *3*, 1218–1223. doi: 10.1038/81504

D'Cruz, A., Ragozzino, M. E., Mosconi, M. W., Shrestha, S., Cook, E. H., & Sweeney, J. A. (2013). Reduced behavioral flexibility in autism spectrum disorders. *Neuropsychology*, *27*(2), 152–160. doi: 10.1037/a0031721

de-Wit, L. H., Kubilius, J., Wagemans, J., & Beeck, H. P. O. d. (2012). Bistable gestalts reduce activity in the whole of v1, not just the retinotopically predicted parts. *Journal of Vision*, *12*(11), 12. doi: 10.1167/12.11.12

de-Wit, L. H., Machilsen, B., & Putzeys, T. (2010). Predictive coding and the neural response to predictable stimuli. *Journal of Neuroscience*, *30*(26), 8702–8703. doi: 10.1523/JNEUROSCI.2248-10.2010

Deacon, T. W. (2010). What is missing from theories of information? In P. Davies & N. H. Gregersen (Eds.), *Information and the nature of reality* (p. 146). New York: Cambridge University Press.

Deacon, T. W. (2011). *Incomplete nature: How mind emerged from matter*. W. W. Norton & Company.

de Gardelle, V., Waszczuk, M., Egner, T., & Summerfield, C. (2012). Concurrent repetition enhancement and suppression responses in extrastriate visual cortex. *Cerebral Cortex*. doi: 10.1093/cercor/bhs211

Dehaene, S., & Cohen, L. (2007). Cultural recycling of cortical maps. *Neuron*, *56*(2), 384–398. doi: 10.1016/j.neuron.2007.10.004

De Houwer, J., Hermans, D., Rothermund, K., & Wentura, D. (2002). Affective priming of semantic categorisation responses. *Cognition & Emotion*, *16*(5), 643–666. doi: 10.1080/02699930143000419

de Jonge, M. V., Kemner, C., De Haan, E. H., Coppens, J. E., Van

Den Berg, T. J. T. P., & Van Engeland, H. (2007). Visual information processing in high-functioning individuals with autism spectrum disorders and their parents. *Neuropsychology*, *21*(1), 65–73. doi: 10.1037/0894-4105.21.1.65

Delorme, R., Ey, E., Toro, R., Leboyer, M., Gillberg, C., & Bourgeron, T. (2013). Progress toward treatments for synaptic defects in autism. *Nature Medicine*, *19*(6), 685–694. doi: 10.1038/nm.3193

Del Viva, M. M., Igliozzi, R., Tancredi, R., & Brizzolara, D. (2006). Spatial and motion integration in children with autism. *Vision Research*, *46*(8–9), 1242–1252. doi: 10.1016/j.visres.2005.10.018

Dember, W. N., Earl, R. W., & Paradise, N. (1957). Response by rats to differential stimulus complexity. *Journal of Comparative and Physiological Psychology*, *50*(5), 514–518. doi: 10.1037/h0046780

den Ouden, H. E. M., Daunizeau, J., Roiser, J., Friston, K. J., & Stephan, K. E. (2010). Striatal prediction error modulates cortical coupling. *The Journal of Neuroscience*, *30*(9), 3210 –3219. doi: 10.1523/JNEUROSCI.4458-09.2010

den Ouden, H. E. M., Friston, K., Daw, N., McIntosh, A., & Stephan, K. (2009). A dual role for prediction error in associative learning. *Cerebral Cortex*, *19*(1460-2199 (Electronic)), 1175–1185. doi: 10.1093/cercor/bhn161

den Ouden, H. E. M., Kok, P., & Lange, F. P. d. (2012). How prediction errors shape perception, attention, and motivation. *Frontiers in Perception Science*, *3*, 548. doi: 10.3389/fpsyg.2012.00548

Dewey, J. (2005). *Art as experience*. Penguin.

Dichter, G. S., Felder, J. N., & Bodfish, J. W. (2009). Autism is characterized by dorsal anterior cingulate hyperactivation during social target detection. *Social Cognitive and Affective Neuroscience*, *4*(3), 215–226. doi: 10.1093/scan/nsp017

Dijksterhuis, A., & Aarts, H. (2003). On wildebeests and humans. *Psychological Science*, *14*(1), 14 –18. doi: 10.1111/1467-9280.t01-1-01412

Di Martino, A., Ross, K., Uddin, L. Q., Sklar, A. B., Castellanos, F. X., & Milham, M. P. (2009). Functional brain correlates of social and nonsocial processes in autism spectrum disorders: an activation likelihood estimation meta-analysis. *Biological psychiatry*, *65*(1), 63–74. doi: 10.1016/j.biopsych.2008.09.022

Di Paolo, E. A. (2003). Organismically-inspired robotics: homeostatic adaptation and teleology beyond the closed sensorimotor loop. *Dynamical systems approach to embodiment and sociality*, 19–42.

Dissanayake, E. (2008). The arts after darwin: Does art have an origin and adaptive function. *World art studies: Exploring concepts*

and approaches, 241–263.

Dissanayake, E. (2009). The artification hypothesis and its relevance to cognitive science, evolutionary aesthetics, and neuroaesthetics. *Cognitive Semiotics*, *9*(5), 136–158.

Diuk, C., Tsai, K., Wallis, J., Botvinick, M., & Niv, Y. (2013). Hierarchical learning induces two simultaneous, but separable, prediction errors in human basal ganglia. *The Journal of Neuroscience*, *33*(13), 5797–5805. doi: 10.1523/JNEUROSCI.5445-12.2013

Dodgson, N. (2009). Balancing the expected and the surprising in geometric patterns. *Computers & Graphics*, *33*(4), 475–483. doi: 10.1016/j.cag.2009.04.001

Dosenbach, N. U., Visscher, K. M., Palmer, E. D., Miezin, F. M., Wenger, K. K., Kang, H. C., ... Petersen, S. E. (2006). A core system for the implementation of task sets. *Neuron*, *50*(5), 799–812. doi: 10.1016/j.neuron.2006.04.031

Dreisbach, G., & Fischer, R. (2012). Conflicts as aversive signals. *Brain & Cognition*, *78*(2), 94–98. doi: 10.1016/j.bandc.2011.12.003

Duncan, S., & Barrett, L. F. (2007). Affect is a form of cognition: A neurobiological analysis. *Cognition & emotion*, *21*(6), 1184–1211. doi: 10.1080/02699930701437931

Dunn, M., Gomes, H., & Gravel, J. (2008). Mismatch negativity in children with autism and typical development. *Journal of Autism and Developmental Disorders*, *38*(0162-3257 (Print)), 52–71. doi: 10.1007/s10803-007-0359-3

Dunning, D., & Balcetis, E. (2013). Wishful seeing: How preferences shape visual perception. *Current Directions in Psychological Science*, *22*(1), 33–37. doi: 10.1177/0963721412463693

Dutton, D. G., & Aron, A. P. (1974). Some evidence for heightened sexual attraction under conditions of high anxiety. *Journal of Personality and Social Psychology*, *30*(4), 510–517. doi: 10.1037/h0037031

Duzel, E., & Guitart-Masip, M. (2013). Not so uncertain at last: Locus coeruleus and decision making. *Neuron*, *79*(1), 9–11. doi: 10.1016/j.neuron.2013.06.023

Ebert, D. H., & Greenberg, M. E. (2013). Activity-dependent neuronal signalling and autism spectrum disorder. *Nature*, *493*(7432), 327–337. doi: 10.1038/nature11860

Edelson, S. M., Edelson, M. G., Kerr, D. C. R., & Grandin, T. (1999). Behavioral and physiological effects of deep pressure on children with autism: A pilot study evaluating the efficacy of grandin's hug machine. *American Journal of Occupational Therapy*, *53*(2), 145–152. doi: 10.5014/ajot.53.2.145

Edwards, D. J., Perlman, A., & Reed, P. (2012). Unsupervised categorization in a sample of children with autism spectrum disorders. *Research in Developmental Disabilities*, *33*(4), 1264–1269. doi: 10.1016/j.ridd.2012.02.021

Edwards, M. J., Adams, R. A., Brown, H., Pareés, I., & Friston, K. J. (2012). A Bayesian account of 'hysteria'. *Brain*, *135*(11), 3495–3512. doi: 10.1093/brain/aws129

Egbert, M. (2013). For biological systems, maintaining essential variables within viability limits is not passive. *Constructivist Foundations*, *9*(1), 109–111.

Egbert, M., & Barandiaran, X. E. (2014). Modeling habits as self-sustaining patterns of sensorimotor behavior. *Frontiers in Human Neuroscience*, *8*. doi: 10.3389/fnhum.2014.00590

Egbert, M., & Canamero, L. (2014). Habit-based regulation of essential variables. In H. Sayama, J. Rieffel, S. Risi, R. Doursat, & H. Lipson (Eds.), *Artificial life 14: Proceedings of the fourteenth international conference on the synthesis and simulation of living systems.* Cambridge, MA: MIT Press.

Egner, T., Monti, J. M., & Summerfield, C. (2010). Expectation and surprise determine neural population responses in the ventral visual stream. *The Journal of Neuroscience*, *30*(49), 16601 –16608. doi: 10.1523/JNEUROSCI.2770-10.2010

Egner, T., & Summerfield, C. (2013). Grounding predictive coding models in empirical neuroscience research. *Behavioral and Brain Sciences*, *36*(03), 210–211. doi: 10.1017/S0140525X1200218X

Elison, J. T., Paterson, S. J., Wolff, J. J., Reznick, J. S., Sasson, N. J., Gu, H., ... Piven, J. (2013). White matter microstructure and atypical visual orienting in 7-Month-Olds at risk for autism. *The American journal of psychiatry*. doi: 10.1176/appi.ajp.2012.12091150

Elliott, R., Friston, K. J., & Dolan, R. J. (2000). Dissociable neural responses in human reward systems. *The Journal of Neuroscience*, *20*(16), 6159–6165.

Ellsworth, P. C., & Scherer, K. R. (2003). Appraisal processes in emotion. *Handbook of affective sciences*, *572*.

Elsabbagh, M., Divan, G., Koh, Y., Kim, Y. S., Kauchali, S., Marcín, C., ... Fombonne, E. (2012). Global prevalence of autism and other pervasive developmental disorders. *Autism Research*, *5*(3), 160–179. doi: 10.1002/aur.239

Elsabbagh, M., Volein, A., Holmboe, K., Tucker, L., Csibra, G., Baron-Cohen, S., ... Johnson, M. H. (2009). Visual orienting in the early broader autism phenotype: disengagement and facilitation. *Journal of Child Psychology and Psychiatry*, *50*(5), 637–642.

doi: 10.1111/j.1469-7610.2008.02051.x

Evans, D. W., Orr, P. T., Lazar, S. M., Breton, D., Gerard, J., Ledbetter, D. H., ... Batchelder, H. (2012). Human preferences for symmetry: Subjective experience, cognitive conflict and cortical brain activity. *PLoS ONE*, *7*(6), e38966. doi: 10.1371/journal.pone.0038966

Eysenck, H. (1942). The experimental study of the 'good gestalt'–a new approach. *Psychological Review*, *49*(4), 344–364. doi: 10.1037/h0057013

Fairhall, S. L., & Ishai, A. (2008). Neural correlates of object indeterminacy in art compositions. *Consciousness and Cognition*, *17*(3), 923–932. doi: 10.1016/j.concog.2007.07.005

Falck-Ytter, T. (2008). Face inversion effects in autism: a combined looking time and pupillometric study. *Autism Research*, *1*(5), 297–306. doi: 10.1002/aur.45

Fan, X., Miles, J. H., Takahashi, N., & Yao, G. (2009). Abnormal transient pupillary light reflex in individuals with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *39*(11), 1499–1508. doi: 10.1007/s10803-009-0767-7

Fang, F., Kersten, D., & Murray, S. O. (2008). Perceptual grouping and inverse fMRI activity patterns in human visual cortex. *Journal of Vision*, *8*(7), 1–9. doi: 10.1167/8.7.2

Fazio, R. H. (2001). On the automatic activation of associated evaluations: An overview. *Cognition & Emotion*, *15*(2), 115–141. doi: 10.1080/02699930125908

Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, *4*.

Feldman, J. (2003). What is a visual object? *Trends in Cognitive Sciences*, *7*(6), 252–256. doi: 10.1016/S1364-6613(03)00111-6

Feldman, J. (2009). Bayes and the simplicity principle in perception. *Psychological Review*, *116*(4), 875–887. doi: 10.1037/a0017144

Feldman, J. (2013). Tuning your priors to the world. *Topics in Cognitive Science*, *5*(1), 13–34. doi: 10.1111/tops.12003

Ferri, R., Elia, M., Agarwal, N., Lanuzza, B., Musumeci, S. A., & Pennisi, G. (2003). The mismatch negativity and the p3a components of the auditory event-related potentials in autistic low-functioning subjects. *Clinical Neurophysiology*, *114*(9), 1671–1680. doi: 10.1016/S1388-2457(03)00153-6

Festinger, L. (1962). Cognitive dissonance. *Scientific American*, *207*(4), 93–107. doi: 10.1038/scientificamerican1062-93

Fiorillo, C. D. (2010). A neurocentric approach to Bayesian inference. *Nature Reviews. Neuroscience*. doi: 10.1038/nrn2787-c1

FitzGerald, T. H. B., Dolan, R. J., & Friston, K. J. (2014). Model averaging, optimal inference, and habit formation. *Frontiers in*

Human Neuroscience, *8*, 457. doi: 10.3389/fnhum.2014.00457

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, *76*(5), 378–382. doi: 10.1037/h0031619

Fletcher, P. C., & Frith, C. D. (2009). Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nature Reviews Neuroscience*, *10*(1), 48–58. doi: 10.1038/nrn2536

Foa, E. B., Zinbarg, R., & Rothbaum, B. O. (1992). Uncontrollability and unpredictability in post-traumatic stress disorder: An animal model. *Psychological Bulletin*, *112*(2), 218–238. doi: 10.1037/0033-2909.112.2.218

Foss-Feig, J. H., Tadin, D., Schauder, K. B., & Cascio, C. J. (2013). A substantial and unexpected enhancement of motion perception in autism. *The Journal of Neuroscience*, *33*(19), 8243–8249. doi: 10.1523/JNEUROSCI.1608-12.2013

Fournier, K. A., Hass, C. J., Naik, S. K., Lodha, N., & Cauraugh, J. H. (2010). Motor coordination in autism spectrum disorders: A synthesis and Meta-Analysis. *Journal of Autism and Developmental Disorders*, *40*(10), 1227–1240. doi: 10.1007/s10803-010-0981-3

Freddolino, P. L., & Tavazoie, S. (2012). Beyond homeostasis: a predictive-dynamic framework for understanding cellular behavior. *Annual Review of Cell and Developmental Biology*, *28*, 363–384. doi: 10.1146/annurev-cellbio-092910-154129

Fresco, D. M., Coles, M. E., Heimberg, R. G., Liebowitz, M. R., Hami, S., Stein, M. B., & Goetz, D. (2001). The Liebowitz social anxiety scale: a comparison of the psychometric properties of self-report and clinician-administered formats. *Psychological Medicine*, *31*(06), 1025–1035. doi: 10.1017/S0033291701004056

Frijda, N. H. (1987). Emotion, cognitive structure, and action tendency. *Cognition & Emotion*, *1*, 115–143. doi: 10.1080/02699938708408043

Frijda, N. H. (1988). The laws of emotion. *American Psychologist*, *43*(5), 349–358. doi: 10.1037/0003-066X.43.5.349

Frijda, N. H. (2006). *The laws of emotion* (1edition ed.). Mahwah, N.J: Psychology Press.

Frijda, N. H. (2013). Comment: The why, when, and how of appraisal. *Emotion Review*, *5*(2), 169–170. doi: 10.1177/1754073912468905

Friston, K. J. (2003). Learning and inference in the brain. *Neural Networks*, *16*(9), 1325–1352. doi: 10.1016/j.neunet.2003.06.005

Friston, K. J. (2009). The free-energy principle: a rough guide to the brain? *Trends in Cognitive Sciences*, *13*(7), 293–301. doi: 10.1016/

j.tics.2009.04.005

Friston, K. J. (2010a). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. doi: 10.1038/nrn2787

Friston, K. J. (2010b). Is the free-energy principle neurocentric? *Nature Reviews Neuroscience*. doi: 10.1038/nrn2787-c2

Friston, K. J., Adams, R. A., Perrinet, L., & Breakspear, M. (2012). Perceptions as hypotheses: Saccades as experiments. *Frontiers in Psychology*, *3*. doi: 10.3389/fpsyg.2012.00151

Friston, K. J., Daunizeau, J., & Kiebel, S. J. (2009). Reinforcement learning or active inference? *PLoS ONE*, *4*(7), e6421. doi: 10.1371/journal.pone.0006421

Friston, K. J., Daunizeau, J., Kilner, J., & Kiebel, S. J. (2010). Action and behavior: a free-energy formulation. *Biological Cybernetics*, *102*(3), 227–260. doi: 10.1007/s00422-010-0364-z

Friston, K. J., Lawson, R., & Frith, C. D. (2013). On hyperpriors and hypopriors: comment on pellicano and burr. *Trends in Cognitive Sciences*, *17*(1), 1. doi: 10.1016/j.tics.2012.11.003

Friston, K. J., Schwartenbeck, P., Fitzgerald, T., Moutoussis, M., Behrens, T., & Dolan, R. J. (2013). The anatomy of choice: active inference and agency. *Frontiers in Human Neuroscience*, *7*, 598. doi: 10.3389/fnhum.2013.00598

Friston, K. J., Shiner, T., FitzGerald, T., Galea, J. M., Adams, R., Brown, H., ... Bestmann, S. (2012). Dopamine, affordance and active inference. *PLoS computational biology*, *8*(1), e1002327. doi: 10.1371/journal.pcbi.1002327

Friston, K. J., Thornton, C., & Clark, A. (2012). Free-Energy minimization and the Dark-Room problem. *Frontiers in Psychology*, *3*. doi: 10.3389/fpsyg.2012.00130

Frith, U. (2003). *Autism: Explaining the enigma (2nd ed.)*. Oxford: Blackwell.

Frith, U., & Happé, F. (1994). Autism: beyond "theory of mind". *Cognition*, *50*(1-3), 115–132.

Froese, T., & Ikegami, T. (2013). The brain is not an isolated "black box," nor is its goal to become one. *Behavioral and Brain Sciences*, *36*(03), 213–214. doi: 10.1017/S0140525X12002348

Furmark, T., Tillfors, M., Everz, P., Marteinsdottir, I., Gefvert, O., & Fredrikson, M. (1999). Social phobia in the general population: prevalence and sociodemographic profile. *Social Psychiatry and Psychiatric Epidemiology*, *34*(8), 416–424. doi: 10.1007/s001270050163

Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2012). *Irr: various coefficients of interrater reliability and agreement (R package version 0.83)*.

Gastgeb, H., Dundas, E. M., Minshew, N. J., & Strauss, M. S. (2012). Category formation in autism: Can individuals with autism form categories and prototypes of dot patterns? *Journal of Autism and Developmental Disorders*, *42*(8), 1694–1704. doi: 10.1007/s10803-011-1411-x

Gastgeb, H., Rump, K. M., Best, C. A., Minshew, N. J., & Strauss, M. S. (2009). Prototype formation in autism: Can individuals with autism abstract facial prototypes? *Autism Research*, *2*(5), 279–284. doi: 10.1002/aur.93

Gastgeb, H., Wilkinson, D., Minshew, N., & Strauss, M. (2011). Can individuals with autism abstract prototypes of natural faces? *Journal of Autism and Developmental Disorders*, *41*(12), 1609–1618. doi: 10.1007/s10803-011-1190-4

Gaver, W., & Mandler, G. (1987). Play it again, sam: On liking music. *Cognition & Emotion*, *1*(3), 259–282. doi: 10.1080/02699938708408051

Gentili, C., Gobbini, M. I., Ricciardi, E., Vanello, N., Pietrini, P., Haxby, J. V., & Guazzelli, M. (2008). Differential modulation of neural activity throughout the distributed neural system for face perception in patients with social phobia and healthy subjects. *Brain Research Bulletin*, *77*(5), 286–292.

Gerken, L., Balcomb, F. K., & Minton, J. L. (2011). Infants avoid 'labouring in vain' by attending more to learnable than unlearnable linguistic patterns. *Developmental Science*, *14*(5), 972–979. doi: 10.1111/j.1467-7687.2011.01046.x

Gershman, S. J., & Daw, N. D. (2012). Perception, action and utility: The tangled skein. *Principles of brain dynamics: Global state interactions*, 293–312.

Geschwind, D. (2011). Genetics of autism spectrum disorders. *Trends in Cognitive Sciences*, *15*(1879-307X (Electronic)), 409–416. doi: 10.1016/j.tics.2011.07.003

Geurts, H. M., Corbett, B., & Solomon, M. (2009). The paradox of cognitive flexibility in autism. *Trends in Cognitive Sciences*, *13*(2), 74–82. doi: 10.1016/j.tics.2008.11.006

Gibson, E. (1969). *Principles of perceptual learning and development*. East Norwalk, CT, US: Appleton-Century-Crofts.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton Mifflin.

Gioia, G. A., Isquith, P. K., Kenworthy, L., & Barton, R. M. (2002). Profiles of everyday executive function in acquired and developmental disorders. *Child Neuropsychology*, *8*(2), 121–137. doi: 10.1076/chin.8.2.121.8727

Gomot, M., Blanc, R., Clery, H., Roux, S., Barthelemy, C., & Bruneau, N. (2011). Candidate electrophysiological endophenotypes of

hyper-reactivity to change in autism. *Journal of Autism and Developmental Disorders*, *41*(1573-3432 (Electronic)), 705–714. doi: 10.1007/s10803-010-1091-y

Gomot, M., Giard, M., Adrien, J., Barthelemy, C., & Bruneau, N. (2002). Hypersensitivity to acoustic change in children with autism: electrophysiological evidence of left frontal corex dysfunctioning. *Psychophysiology*, *39*(0048-5772 (Print)), 577–584.

Gomot, M., & Wicker, B. (2012). A challenging, unpredictable world for people with autism spectrum disorder. *International Journal of Psychophysiology*, *83*(2), 240–247. doi: 10.1016/j.ijpsycho.2011.09.017

Gonzalez, C., Martin, J. M., Minshew, N. J., & Behrmann, M. (2013). Practice makes improvement: How adults with autism Out-Perform others in a naturalistic visual search task. *Journal of Autism and Developmental Disorders*, 1–10. doi: 10.1007/s10803-013-1772-4

Goris, R. L. T., Wagemans, J., & Wichmann, F. A. (2008). Modelling contrast discrimination data suggest both the pedestal effect and stochastic resonance to be caused by the same mechanism. *Journal of Vision*, *8*(15). doi: 10.1167/8.15.17

Gorlin, S., Meng, M., Sharma, J., Sugihara, H., Sur, M., & Sinha, P. (2012). Imaging prior information in the brain. *Proceedings of the National Academy of Sciences*, *109*(20), 7935–7940. doi: 10.1073/pnas.1111224109

Gottlieb, J. (2012). Attention, learning, and the value of information. *Neuron*, *76*(2), 281–295. doi: 10.1016/j.neuron.2012.09.034

Gradin, V. B., Kumar, P., Waiter, G., Ahearn, T., Stickle, C., Milders, M., ... Steele, J. D. (2011). Expected value and prediction error abnormalities in depression and schizophrenia. *Brain*, *134*(6), 1751–1764. doi: 10.1093/brain/awr059

Grandin, T. (1992). Calming effects of deep touch pressure in patients with autistic disorder, college students, and animals. *Journal of Child & Adolescent Psychopharmacology*, *2*(1044-5463 (Print)), 63–72. doi: 10.1089/cap.1992.2.63

Grandin, T. (1995). *Thinking in pictures and other reports of my life with autism*. New York: Vintage Books Inc.

Gray, K. L. H., Adams, W. J., & Garner, M. (2009). The influence of anxiety on the initial selection of emotional faces presented in binocular rivalry. *Cognition*, *113*(1), 105–110. doi: 10.1016/j.cognition.2009.06.009

Greenaway, K. H., Louis, W. R., & Hornsey, M. J. (2013). Loss of control increases belief in precognition and belief in precognition increases control. *PLoS ONE*, *8*(8). doi: 10.1371/journal.pone.0071327

Gregory, R. L. (1980). Perceptions as hypotheses. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *290*(1038), 181–197. doi: 10.1098/rstb.1980.0090

Griffiths, T. L., Chater, N., Norris, D., & Pouget, A. (2012). How the Bayesians got their beliefs (and what those beliefs actually are): Comment on Bowers and Davis (2012). *Psychological Bulletin*, *138*(3), 415–422. doi: 10.1037/a0026884

Gu, X., Hof, P. R., Friston, K. J., & Fan, J. (2013). Anterior insular cortex and emotional awareness. *Journal of Comparative Neurology*, *521*(15), 3371–3388. doi: 10.1002/cne.23368

Haesen, B., Boets, B., & Wagemans, J. (2011). A review of behavioural and electrophysiological studies on auditory processing and speech perception in autism spectrum disorders. *Research in Autism Spectrum Disorders*, *5*, 701–714.

Ham, T., Leff, A., Boissezon, X. d., Joffe, A., & Sharp, D. J. (2013). Cognitive control and the salience network: An investigation of error processing and effective connectivity. *The Journal of Neuroscience*, *33*(16), 7091–7098. doi: 10.1523/JNEUROSCI.4692-12.2013

Hamilton, A. F. d. C., Brindley, R. M., & Frith, U. (2007). Imitation and action understanding in autistic spectrum disorders: How valid is the hypothesis of a deficit in the mirror neuron system? *Neuropsychologia*, *45*(8), 1859–1868. doi: 10.1016/j.neuropsychologia.2006.11.022

Happé, F., & Frith, U. (2006). The weak coherence account: Detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *36*(1), 5–25. doi: 10.1007/s10803-005-0039-0

Happe, F., & Ronald, A. (2008). The 'fractionable autism triad': a review of evidence from behavioural, genetic, cognitive and neural research. *Neuropsychology Review*, *18*(1040-7308 (Print)), 287–304. doi: 10.1007/s11065-008-9076-8

Happé, F. G. E., & Booth, R. D. L. (2008). The power of the positive: Revisiting weak coherence in autism spectrum disorders. *The Quarterly Journal of Experimental Psychology*, *61*(1), 50–63. doi: 10.1080/17470210701508731

Harms, M. B., Martin, A., & Wallace, G. L. (2010). Facial emotion recognition in autism spectrum disorders: A review of behavioral and neuroimaging studies. *Neuropsychology Review*, *20*(3), 290–322. doi: 10.1007/s11065-010-9138-6

Hawkins, J., & Blakeslee, S. (2004). *On intelligence* (Adapted ed.). Times Books.

Heaton, P. (2003). Pitch memory, labelling and disembedding in autism. *Journal of Child Psychology and Psychiatry*, *44*(4), 543–

551. doi: 10.1111/1469-7610.00143

Heaton, P., Hudry, K., Ludlow, A., & Hill, E. (2008). Superior discrimination of speech pitch and its relationship to verbal ability in autism spectrum disorders. *Cognitive Neuropsychology*, *25*(1464-0627 (Electronic)), 771–782. doi: 10.1080/02643290802336277

Heaton, P., Williams, K., Cummins, O., & Happe, F. (2008). Autism and pitch processing splinter skills: a group and subgroup analysis. *Autism*, *12*(1362-3613 (Print)), 203–219. doi: 10.1177/1362361307085270

Hebb, D. (1946). On the nature of fear. *Psychological Review*, *53*(5), 259–276. doi: 10.1037/h0061690

Hebb, D. (1949). *The organization of behavior*. New York: John Wiley.

Heenan, A., & Troje, N. F. (2014). Both physical exercise and progressive muscle relaxation reduce the Facing-the-Viewer bias in biological motion perception. *PLoS ONE*, *9*(7), e99902. doi: 10.1371/journal.pone.0099902

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology*, *57*(2), 243. doi: 10.2307/1416950

Heine, S. J., Proulx, T., & Vohs, K. D. (2006). The meaning maintenance model: On the coherence of social motivations. *Personality and Social Psychology Review*, *10*(2), 88–110. doi: 10.1207/s15327957pspr1002_1

Helmholtz, H. v. (1878). The facts of perception. In R. Kahl (Ed.), *Selected writings of hermann helmholtz.* Wesleyan University Press.

Helmholtz, H. v. (1962). *Treatise on physiological optics*. New York: Dover.

Herbart, J. F. (1891). *A text-book in psychology: An attempt to found the science of psychology on experience, metaphysics, and mathematics* (Vol. 18). D. Appleton.

Hermans, D., & De Houwer, J. (1994). Affective and subjective familiarity ratings of 740 dutch words. *Psychologica Belgica*, *34*(2-3), 115–139.

Hermans, D., De Houwer, J., & Eelen, P. (2001). A time course analysis of the affective priming effect. *Cognition & Emotion*, *15*(2), 143–165. doi: 10.1080/02699930125768

Hermans, D., Spruyt, A., De Houwer, J., & Eelen, P. (2003). Affective priming with subliminally presented pictures. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *57*(2), 97–114. doi: 10.1037/h0087416

Herry, C., Bach, D. R., Esposito, F., Di Salle, F., Perrig, W. J., Scheffler, K., ... Seifritz, E. (2007). Processing of temporal unpredictability in human and animal amygdala. *Journal of Neuroscience*,

27(22), 5958–5966. doi: 10.1523/JNEUROSCI.5218-06.2007

Heylighen, F., & Joslyn, C. (2001). Cybernetics and second order cybernetics. *Encyclopedia of physical science & technology*, *4*, 155–170.

Hill, A. L. (1978). Savants: Mentally retarded individuals with special skills. *International Review of Research in Mental Retardation*, *9*, 277–298.

Hill, E. L. (2004). Executive dysfunction in autism. *Trends in Cognitive Sciences*, *8*(1), 26–32. doi: 10.1016/j.tics.2003.11.003

Hillyard, S. A., Vogel, E. K., & Luck, S. J. (1998). Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *353*(1373), 1257–1270. doi: 10.1098/rstb.1998.0281

Hirsh, J. B., Mar, R. A., & Peterson, J. B. (2012). Psychological entropy: a framework for understanding uncertainty-related anxiety. *Psychological Review*, *119*(2), 304–320. doi: 10.1037/a0026767

Hoffman, D. (2009). The interface theory of perception: Natural selection drives true perception to swift extinction. In S. Dickinson, M. Tarr, A. Leonardis, & B. Schiele (Eds.), *Object categorization: Computer and human vision perspectives* (pp. 148–165). Cambridge, U.K.: Cambridge University Press.

Hoffmann, J. (2003). Anticipatory behavioral control. In M. V. Butz, O. Sigaud, & P. Gérard (Eds.), *Anticipatory behavior in adaptive learning systems* (pp. 44–65). Springer Berlin Heidelberg.

Hoge, E. A., Pollack, M. H., Kaufman, R. E., Zak, P. J., & Simon, N. M. (2008). Oxytocin levels in social anxiety disorder. *CNS Neuroscience & Therapeutics*, *14*(3), 165–170. doi: 10.1111/j.1755-5949.2008.00051.x

Hohwy, J. (2011). Phenomenal variability and introspective reliability. *Mind & Language*, *26*(3), 261–286. doi: 10.1111/j.1468-0017.2011.01418.x

Hohwy, J. (2013). *The predictive mind*. Oxford University Press.

Hohwy, J. (2014). The Self-Evidencing brain. *Noûs*. doi: 10.1111/nous.12062

Hohwy, J., & Palmer, C. (2014). Social cognition as causal inference: implications for common knowledge and autism. In M. Gallotti & J. Michael (Eds.), *Social ontology and social cognition (forthcoming)*.

Hohwy, J., Roepstorff, A., & Friston, K. (2008). Predictive coding explains binocular rivalry: An epistemological review. *Cognition*, *108*(3), 687–701. doi: 10.1016/j.cognition.2008.05.010

Honey, R. C., Watt, A., & Good, M. (1998). Hippocampal lesions

disrupt an associative mismatch process. *The Journal of Neuroscience*, *18*(6), 2226–2230.

Horn, B. K. P. (1980). Derivation of invariant scene characteristics from images. In *Proceedings of the may 19-22, 1980, national computer conference* (p. 371–376). New York, NY, USA: ACM. doi: 10.1145/1500518.1500579

Howlin, P., Goode, S., Hutton, J., & Rutter, M. (2009). Savant skills in autism: psychometric approaches and parental reports. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1522), 1359–1367. doi: 10.1098/rstb.2008.0328

Hsieh, P., Vul, E., & Kanwisher, N. (2010). Recognition alters the spatial pattern of fMRI activation in early retinotopic cortex. *Journal of Neurophysiology*, *103*(3), 1501–1507. doi: 10.1152/jn.00812.2009

Huang, Y., & Rao, R. P. N. (2011). Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(5), 580–593. doi: 10.1002/wcs.142

Hughes, C. (1996). Brief report: Planning problems in autism at the level of motor control. *Journal of Autism and Developmental Disorders*, *26*(1), 99–107. doi: 10.1007/BF02276237

Hurley, M. M., Dennett, D. C., & Adams, J., Reginald B. (2011). *Inside jokes: Using humor to Reverse-Engineer the mind*. MIT Press.

Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation* (1st ed.). The MIT Press.

Hutsler, J. J., & Zhang, H. (2010). Increased dendritic spine densities on cortical projection neurons in autism spectrum disorders. *Brain Research*, *1309*, 83–94. doi: 10.1016/j.brainres.2009.09.120

Huys, Q. J. M., & Dayan, P. (2009). A Bayesian formulation of behavioral control. *Cognition*, *113*(3), 314–328. doi: 10.1016/j.cognition.2009.01.008

Hyman, J. (2010). Art and neuroscience. In R. Frigg & M. Hunter (Eds.), *Beyond mimesis and convention* (Vol. 262, pp. 245–261). Dordrecht: Springer Netherlands.

Iarocci, G., Burack, J. A., Shore, D. I., Mottron, L., & Enns, J. T. (2006). Global–Local visual processing in high functioning children with autism: Structural vs. implicit task biases. *Journal of Autism and Developmental Disorders*, *36*(1), 117–129. doi: 10.1007/s10803-005-0045-2

Ide, J. S., Shenoy, P., Yu, A. J., & Li, C.-s. R. (2013). Bayesian prediction and evaluation in the anterior cingulate cortex. *The Journal of Neuroscience*, *33*(5), 2039–2047. doi: 10.1523/JNEUROSCI.2201-12.2013

Iglesias, S., Mathys, C., Brodersen, K. H., Kasper, L., Piccirelli, M., den Ouden, H. E. M., & Stephan, K. E. (2013). Hierarchi-

cal prediction errors in midbrain and basal forebrain during sensory learning. *Neuron*, *80*(2), 519–530. doi: 10.1016/j.neuron.2013.09.009

Imamoglu, F., Kahnt, T., Koch, C., & Haynes, J. (2012). Changes in functional connectivity support conscious object recognition. *NeuroImage*, *63*(4), 1909–1917. doi: 10.1016/j.neuroimage.2012.07.056

Inoue, W., Baimoukhametova, D. V., Füzesi, T., Cusulin, J. I. W., Koblinger, K., Whelan, P. J., ... Bains, J. S. (2013). Noradrenaline is a stress-associated metaplastic signal at GABA synapses. *Nature Neuroscience*, *16*(5), 605–612. doi: 10.1038/nn.3373

Inzlicht, M., Tullett, A. M., & Good, M. (2011). The need to believe: a neuroscience account of religion as a motivated process. *Religion, Brain & Behavior*, *1*(3), 192–212. doi: 10.1080/2153599X.2011.647849

Ishai, A., Fairhall, S. L., & Pepperell, R. (2007). Perception, memory and aesthetics of indeterminate art. *Brain Research Bulletin*, *73*(4-6), 319–324. doi: 10.1016/j.brainresbull.2007.04.009

Izard, C. E. (2007). Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspectives on Psychological Science*, *2*(3), 260–280. doi: 10.1111/j.1745-6916.2007.00044.x

Jackendoff, R., & Lerdahl, F. (2006). The capacity for music: What is it, and what's special about it? *Cognition*, *100*(1), 33–72. doi: 10.1016/j.cognition.2005.11.005

Jackson, F., Nelson, B. D., & Proudfit, G. H. (2014). In an uncertain world, errors are more aversive: Evidence from the Error-Related negativity. *Emotion*. doi: 10.1037/emo0000020

James, W. (1890). The principles of psychology. *Harvard UP, Cambridge, MA*.

James, W. (2000). *Pragmatism (1907)*. New York: Penguin Books.

Jarvinen-Pasley, A., Wallace, G., Ramus, F., Happe, F., & Heaton, P. (2008). Enhanced perceptual processing of speech in autism. *Developmental Science*, *11*(1467-7687 (Electronic)), 109–121. doi: 10.1111/j.1467-7687.2007.00644.x

Jay, M. (1994). *Downcast eyes: The denigration of vision in Twentieth-Century french thought* (1ST edition ed.). Berkeley: University of California Press.

Jepma, M., Verdonschot, R. G., van Steenbergen, H., Rombouts, S. A. R. B., & Nieuwenhuis, S. (2012). Neural mechanisms underlying the induction and relief of perceptual curiosity. *Frontiers in Behavioral Neuroscience*, *6*. doi: 10.3389/fnbeh.2012.00005

Joffily, M., & Coricelli, G. (2013). Emotional valence and the Free-

Energy principle. *PLoS Computational Biology*, *9*(6), e1003094. doi: 10.1371/journal.pcbi.1003094

Johnson, A. W., & Gallagher, M. (2010). Greater effort boosts the affective taste properties of food. *Proceedings of the Royal Society B: Biological Sciences*. doi: 10.1098/rspb.2010.1581

Jones, C., Happe, F., Baird, G., Simonoff, E., Marsden, A., Tregay, J., ... Charman, T. (2009). Auditory discrimination and auditory sensory behaviours in autism spectrum disorders. *Neuropsychologia*, *47*(1873-3514 (Electronic)), 2850–2858. doi: 10.1016/j.neuropsychologia.2009.06.015

Joseph, R. M., Keehn, B., Connolly, C., Wolfe, J. M., & Horowitz, T. S. (2009). Why is visual search superior in autism spectrum disorder? *Developmental Science*, *12*(6), 1083–1096. doi: 10.1111/j.1467-7687.2009.00855.x

Kagan, J. (1972). Motives and development. *Journal of Personality and Social Psychology*, *22*(1), 51–66. doi: 10.1037/h0032356

Kagan, J. (2002). *Surprise, uncertainty, and mental structures*. Harvard University Press.

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child*, *2*, 217–250.

Kaplan, F., & Oudeyer, P. (2007). In search of the neural circuits of intrinsic motivation. *Frontiers in Neuroscience*, *1*(1), 225–236. doi: 10.3389/neuro.01.1.1.017.2007

Kaplan, H. R. (1987). Lottery winners: The myth and reality. *Journal of gambling behavior*, *3*(3), 168–178. doi: 10.1007/BF01367438

Kapp, B. S., Whalen, P. J., Supple, W. F., & Pascoe, J. P. (1992). Amygdaloid contributions to conditioned arousal and sensory information processing. In *The amygdala: Neurobiological aspects of emotion, memory, and mental dysfunction* (pp. 229–254). New York, NY, US: Wiley-Liss.

Karatsoreos, I. N., & McEwen, B. S. (2011). Psychobiological allostasis: resistance, resilience and vulnerability. *Trends in Cognitive Sciences*, *15*(12), 576–584. doi: 10.1016/j.tics.2011.10.005

Karlsson, M. P., Tervo, D. G. R., & Karpova, A. Y. (2012). Network resets in medial prefrontal cortex mark the onset of behavioral uncertainty. *Science*, *338*(6103), 135–139. doi: 10.1126/science.1226518

Karmiloff-Smith, A. (1994). Précis of beyond modularity: A developmental perspective on cognitive science. *Behavioral and Brain Sciences*, *17*(04), 693–707. doi: 10.1017/S0140525X00036621

Keehn, B., Müller, R., & Townsend, J. (2013). Atypical attentional networks and the emergence of autism. *Neuroscience & Biobehavioral Reviews*, *37*(2), 164–183. doi: 10.1016/j.neubiorev.2012.11.014

Keita, L., Mottron, L., & Bertone, A. (2010). Far visual acuity is unremarkable in autism: do we need to focus on crowding? *Autism Research*, *3*(1939-3806 (Electronic)), 333–341. doi: 10.1002/aur.164

Kemner, C., Verbaten, M., Cuperus, J., Camfferman, G., & van Engeland, H. (1995). Auditory event-related brain potentials in autistic children and three different control groups. *Biological Psychiatry*, *38*(0006-3223 (Print)), 150–165. doi: 10.1016/0006-3223(94)00247-Z

Kenworthy, L., Yerys, B. E., Anthony, L. G., & Wallace, G. L. (2008). Understanding executive control in autism spectrum disorders in the lab and in the real world. *Neuropsychology Review*, *18*(4), 320–338. doi: 10.1007/s11065-008-9077-7

Keramati, M., & Gutkin, B. S. (2011). A reinforcement learning theory for homeostatic regulation. In *NIPS* (p. 82–90).

Kéri, S., & Benedek, G. (2009). Oxytocin enhances the perception of biological motion in humans. *Cognitive, Affective, & Behavioral Neuroscience*, *9*(3), 237–241. doi: 10.3758/CABN.9.3.237

Kern, J. K., Miller, V. S., Cauller, L., Kendall, R., Mehta, J., & Dodd, M. (2001). Effectiveness of N,N-Dimethylglycine in autism and pervasive developmental disorder. *Journal of Child Neurology*, *16*(3), 169–173. doi: 10.1177/088307380101600303

Kern, J. K., Trivedi, M., Garver, C., Grannemann, B., Andrews, A., Savla, J., ... Schroeder, J. (2006). The pattern of sensory processing abnormalities in autism. *Autism*, *10*(1362-3613 (Print)), 480–494. doi: 10.1177/1362361306066564

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology*, *55*, 271–304. doi: 10.1146/annurev.psych.55.090902.142005

Khalfa, S., Bruneau, N., Rogé, B., Georgieff, N., Veuillet, E., Adrien, J., ... Collet, L. (2004). Increased perception of loudness in autism. *Hearing Research*, *198*(1–2), 87–92. doi: 10.1016/j.heares.2004.07.006

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS ONE*, *7*(5), e36399. doi: 10.1371/journal.pone.0036399

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). Predictive coding: an account of the mirror neuron system. *Cognitive Processing*, *8*(3), 159–166. doi: 10.1007/s10339-007-0170-2

Kim, J. A., Szatmari, P., Bryson, S. E., Streiner, D. L., & Wilson, F. J. (2000). The prevalence of anxiety and mood problems among children with autism and asperger syndrome. *Autism*, *4*(2), 117–132. doi: 10.1177/1362361300004002002

Kimura, M., Schröger, E., & Czigler, I. (2011). Visual mismatch negativity and its importance in visual cognitive sciences. *NeuroReport*, *22*, 669–673. doi: 10.1097/WNR.0b013e32834973ba

Klauer, K. C., & Musch, J. (2003). Affective priming: Findings and theories. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 7–49). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Klein, C. (2014). The penumbral theory of masochistic pleasure. *Review of Philosophy and Psychology*, *5*(1), 41–55. doi: 10.1007/s13164-013-0169-9

Klein, D. C., & Seligman, M. E. (1976). Reversal of performance deficits and perceptual deficits in learned helplessness and depression. *Journal of Abnormal Psychology*, *85*(1), 11–26. doi: 10.1037/0021-843X.85.1.11

Klin, A. (1991). Young autistic childrens listening preferences in regard to speech - a possible characterization of the symptom of social withdrawal. *Journal of Autism and Developmental Disorders*, *21*(1), 29–42. doi: 10.1007/BF02206995

Klin, A. (1992). Listening preferences in regard to speech in 4 Children with Developmental-Disabilities. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *33*(4), 763–769. doi: 10.1111/j.1469-7610.1992.tb00911.x

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of General Psychiatry*, *59*, 809–816.

Klin, A., Lin, D., Gorrindo, P., Ramsay, G., & Jones, W. (2009). Two-year-olds with autism orient to non-social contingencies rather than biological motion. *Nature*, *459*(7244), 257–U142. doi: 10.1038/nature07868

Klinger, L. G., & Dawson, G. (2001). Prototype formation in autism. *Development and Psychopathology*, *null*(01), 111–124. doi: null

Knight, E. J., Klepac, K. M., & Kralik, J. D. (2013). Too good to be true: Rhesus monkeys react negatively to Better-than-Expected offers. *PLoS ONE*, *8*(10), e75768. doi: 10.1371/journal.pone.0075768

Koegel, R. L., & Mentis, M. (1985). Motivation in childhood autism: Can they or won't they? *Journal of Child Psychology and Psychiatry*, *26*(2), 185–191. doi: 10.1111/j.1469-7610.1985.tb02259.x

Koenderink, J. (2010). Vision and information. In L. Albertazzi, G. J. Van Tonder, & D. Vishwanath (Eds.), *Perception beyond inference: The information content of visual processes* (p. 27). USA: MIT Press.

Koenderink, J. (2014). The all seeing eye? *Perception*, *43*(1), 1–6.

Koenderink, J., van Doorn, A., & Wagemans, J. (2012). Picasso in the mind's eye of the beholder: Three-dimensional filling-in of ambiguous line drawings. *Cognition*, *125*(3), 394–412. doi: 10.1016/j.cognition.2012.07.019

Koffka, K. (1935). *Principles of gestalt psychology*. New York: Harcourt- Brace.

Kohler, E., Keysers, C., Umiltà, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: Action representation in mirror neurons. *Science*, *297*(5582), 846–848. doi: 10.1126/science.1070311

Köhler, W. (1920). *Die physischen gestalten in ruhe und im stationåren zustand*. Braunschweig, Germany: Vieweg.

Köhler, W. (1938). *The place of value in a world of facts*. New York:: Liveright.

Köhler, W. (1940). *Dynamics in psychology*. New York: Liveright.

Kootz, J. P., & Cohen, D. J. (1981). Modulation of sensory intake in autistic children: Cardiovascular and behavioral indices. *Journal of the American Academy of Child Psychiatry*, *20*(4), 692–701. doi: 10.1097/00004583-198102000-00002

Koster-Hale, J., & Saxe, R. (2013). Theory of mind: A neural prediction problem. *Neuron*, *79*(5), 836–848. doi: 10.1016/j.neuron.2013.08.020

Krupinski, E., & Locher, P. (1988). Skin conductance and aesthetic evaluative responses to nonrepresentational works of art varying in symmetry. *Bulletin of the Psychonomic Society*, *26*(4), 355–358. doi: 10.3758/BF03337681

Kubovy, M. (1999). On the pleasures of the mind. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology.* New York: Russell Sage Foundation.

Kujala, T., Lepisto, T., Nieminen-von Wendt, T., Naatanen, P., & Naatanen, R. (2005). Neurophysiological evidence for cortical discrimination impairment of prosody in asperger syndrome. *Neuroscience Letters*, *383*(0304-3940 (Print)), 260–265. doi: 10.1016/j.neulet.2005.04.048

Kumar, S., Sedley, W., Nourski, K. V., Kawasaki, H., Oya, H., Patterson, R. D., ... Griffiths, T. D. (2011). Predictive coding and pitch processing in the auditory cortex. *Journal of Cognitive Neuroscience*. doi: 10.1162/jocn_a_00021

Lam, K. S., Aman, M. G., & Arnold, L. E. (2006). Neurochemical correlates of autistic disorder: A review of the literature. *Research in Developmental Disabilities*, *27*(3), 254–289. doi: 10.1016/j.ridd.2005.03.003

Landau, M. J., Greenberg, J., Solomon, S., Pyszczynski, T., & Martens, A. (2006). Windows into nothingness: Terror man-

agement, meaninglessness, and negative reactions to modern art. *Journal of Personality and Social Psychology*, *90*(6), 879–892. doi: 10.1037/0022-3514.90.6.879

Lauwereyns, J. (2010). *The anatomy of bias: how neural circuits weigh the options*. MIT Press.

Leder, H., Belke, B., Oeberst, A., & Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology*, *95*(4), 489–508. doi: 10.1348/0007126042369811

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A*, *20*(7), 1434–1448. doi: 10.1364/JOSAA.20.001434

Legare, C. H. (2012). Exploring explanation: Explaining inconsistent evidence informs exploratory, Hypothesis-Testing behavior in young children. *Child Development*, *83*(1), 173–185. doi: 10.1111/j.1467-8624.2011.01691.x

Lepisto, T., Kujala, T., Vanhala, R., Alku, P., Huotilainen, M., & Naatanen, R. (2005). The discrimination of and orienting to speech and non-speech sounds in children with autism. *Brain Research*, *1066*(0006-8993 (Print)), 147–157. doi: 10.1016/j.brainres.2005.10.052

Lepisto, T., Silokallio, S., Nieminen-von Wendt, T., Alku, P., Naatanen, R., & Kujala, T. (2006). Auditory perception and attention as reflected by the brain event-related potentials in children with asperger syndrome. *Clinical Neurophysiology*, *117*(1388-2457 (Print)), 2161–2171. doi: 10.1016/j.clinph.2006.06.709

Lettvin, J. Y., Maturana, H. R., McCulloch, W. S., & Pitts, W. H. (1959). What the frog's eye tells the frog's brain. *Proceedings of the IRE*, *47*(11), 1940–1951.

Liebowitz, M. R. (1987). Social phobia. *Modern Problems of Pharmacopsychiatry*, *22*, 141–173.

Limongi, R., Sutherland, S. C., Zhu, J., Young, M. E., & Habib, R. (2013). Temporal prediction errors modulate cingulate–insular coupling. *NeuroImage*, *71*, 147–157. doi: 10.1016/j.neuroimage.2012.12.078

Lindell, A., & Mueller, J. (2011). Can science account for taste? psychological insights into art appreciation. *Journal of Cognitive Psychology*, *23*(4), 453–475. doi: 10.1080/20445911.2011.539556

Lindquist, K. A. (2013). Emotions emerge from more basic psychological ingredients: A modern psychological constructionist model. *Emotion Review*, *5*(4), 356–368. doi: 10.1177/1754073913489750

Little, D. Y., & Sommer, F. T. (2013). Maximal mutual information, not minimal entropy, for escaping the "Dark room". *Behavioral and Brain Sciences*, *36*(03), 220–221. doi: 10.1017/

S0140525X12002415

López, B., Donnelly, N., Hadwin, J., & Leekam, S. (2004). Face processing in high-functioning adolescents with autism: Evidence for weak central coherence. *Visual Cognition*, *11*(6), 673–688. doi: 10.1080/13506280344000437

Lopez, B. R., Lincoln, A. J., Ozonoff, S., & Lai, Z. (2005). Examining the relationship between executive functions and restricted, repetitive symptoms of autistic disorder. *Journal of Autism and Developmental Disorders*, *35*(4), 445–460. doi: 10.1007/s10803-005-5035-x

Loth, E., Happe, F., & Gomez, J. (2010). Variety is not the spice of life for people with autism spectrum disorders: frequency ratings of central, variable and inappropriate aspects of common real-life events. *Journal of Autism and Developmental Disorders*, *40*(1573-3432 (Electronic)), 730–742. doi: 10.1007/s10803-009-0929-7

Lovaas, O. I. (1987). Behavioral treatment and normal educational and intellectual functioning in young autistic children. *Journal of Consulting and Clinical Psychology*, *55*(1), 3–9.

Lovaas, O. I., Koegel, R. L., & Schreibman, L. (1979). Stimulus overselectivity in autism: a review of research. *Psychological bulletin*, *86*(6), 1236–1254. doi: 10.1037/0033-2909.86.6.1236

Loveland, K. A., Tunali-Kotoski, B., Chen, R., Brelsford, K. A., Ortegon, J., & Pearson, D. A. (1995). Intermodal perception of affect in persons with autism or down syndrome. *Development and Psychopathology*, *7*(03), 409–418. doi: 10.1017/S095457940000660X

Ludmer, R., Dudai, Y., & Rubin, N. (2011). Uncovering camouflage: Amygdala activation predicts Long-Term memory of induced perceptual insight. *Neuron*, *69*(5), 1002–1014. doi: 10.1016/j.neuron.2011.02.013

Ma, W. J., Navalpakkam, V., Beck, J. M., Berg, R. v. d., & Pouget, A. (2011). Behavior and neural basis of near-optimal visual search. *Nature Neuroscience*, *14*(6), 783–790. doi: 10.1038/nn.2814

MacIver, M. A. (2009). Neuroethology: from morphological computation to planning. In P. Robbins & M. Aydede (Eds.), *The cambridge handbook of situated cognition* (p. 480–504). New York, NY: Cambridge University Press.

Mackinlay, R., Charman, T., & Karmiloff-Smith, A. (2006). High functioning children with autism spectrum disorder: A novel test of multitasking. *Brain and Cognition*, *61*(1), 14–24. doi: 10.1016/j.bandc.2005.12.006

Maier, J. X., Neuhoff, J. G., Logothetis, N. K., & Ghazanfar, A. A. (2004). Multisensory integration of looming signals by rhesus

monkeys. *Neuron*, *43*(2), 177–181. doi: 10.1016/j.neuron.2004.06.027

Makin, A. D. J., Pecchinenda, A., & Bertamini, M. (2012). Implicit affective evaluation of visual symmetry. *Emotion*, *12*(5), 1021–1030. doi: 10.1037/a0026924

Mamassian, P. (2008). Ambiguities and conventions in the perception of visual art. *Vision Research*, *48*(20), 2143–2153. doi: 10.1016/j.visres.2008.06.010

Mandler, G. (2003). Emotion. In D. K. Freedheim & I. B. Weiner (Eds.), *History of psychology* (Vol. 1). John Wiley and Sons.

Mark, J. T., Marion, B. B., & Hoffman, D. D. (2010). Natural selection and veridical perceptions. *Journal of Theoretical Biology*, *266*(4), 504–515. doi: 10.1016/j.jtbi.2010.07.020

Markram, K., & Markram, H. (2010). The intense world theory – a unifying theory of the neurobiology of autism. *Frontiers in Human Neuroscience*, *4*. doi: 10.3389/fnhum.2010.00224

Marti, S., Thibault, L., & Dehaene, S. (2014). How does the extraction of local and global auditory regularities vary with context? *PLoS ONE*, *9*(9), e107227. doi: 10.1371/journal.pone.0107227

Mathys, C., Daunizeau, J., Friston, K. J., & Stephan, K. E. (2011). A Bayesian foundation for individual learning under uncertainty. *Frontiers in Human Neuroscience*, *5*. doi: 10.3389/fnhum.2011.00039

Matson, J. L., & Nebel-Schwalm, M. S. (2007). Comorbid psychopathology with autism spectrum disorder in children: an overview. *Research in Developmental Disabilities*, *28*(4), 341–352. doi: 10.1016/j.ridd.2005.12.004

Matson, M. L., Matson, J. L., & Beighley, J. S. (2011). Comorbidity of physical and motor problems in children with autism. *Research in Developmental Disabilities*, *32*(6), 2304–2308. doi: 10.1016/j.ridd.2011.07.036

Maunsell, J. H. R. (2004). Neuronal representations of cognitive state: reward or attention? *Trends in Cognitive Sciences*, *8*(6), 261–265. doi: 10.1016/j.tics.2004.04.003

May, P., & Tiitinen, H. (2010). Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. *Psychophysiology*, *47*(1540-5958 (Electronic)), 66–122. doi: 10.1111/j.1469-8986.2009.00856.x

McNally, G. P., Johansen, J. P., & Blair, H. T. (2011). Placing prediction into the fear circuit. *Trends in Neurosciences*, *34*(6), 283–292. doi: 10.1016/j.tins.2011.03.005

McPartland, J., Dawson, G., Webb, S. J., Panagiotides, H., & Carver, L. J. (2004). Event-related brain potentials reveal anomalies in temporal processing of faces in autism spectrum disorder.

Journal of Child Psychology and Psychiatry, *45*(7), 1235–1245. doi: 10.1111/j.1469-7610.2004.00318.x

Melcher, D. ., & Bacci, F. . (2008). The visual system as a constraint on the survival and success of specific artworks. *Spatial Vision*, *21*, 347–362. doi: 10.1163/156856808784532626

Metereau, E., & Dreher, J. (2013). Cerebral correlates of salient prediction error for different rewards and punishments. *Cerebral Cortex*, *23*(2), 477–487. doi: 10.1093/cercor/bhs037

Meyer. (1961). *Emotion and meaning in music*. University of Chicago Press.

Miall, R. C., Weir, D. J., Wolpert, D. M., & Stein, J. F. (1993). Is the cerebellum a smith predictor? *Journal of Motor Behavior*, *25*(3), 203–216. doi: 10.1080/00222895.1993.9942050

Miller, S. M. (1981). Predictability and human stress: Toward a clarification of evidence and theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. Volume 14, pp. 203–256). Academic Press.

Milne, E., Swettenham, J., Hansen, P., Campbell, R., Jeffries, H., & Plaisted, K. (2002). High motion coherence thresholds in children with autism. *Journal of Child Psychology and Psychiatry*, *43*(2), 255–263. doi: 10.1111/1469-7610.00018

Mineka, S., & Hendersen, R. W. (1985). Controllability and predictability in acquired motivation. *Annual Review of Psychology*, *36*(1), 495–529. doi: 10.1146/annurev.ps.36.020185.002431

Mitchell, P., Mottron, L., Soulieres, I., & Ropar, D. (2010). Susceptibility to the shepard illusion in participants with autism: reduced top-down influences within perception? *Autism Research*, *3*(1939-3806 (Electronic)), 113–119. doi: 10.1002/aur.130

Mobbs, D., Petrovic, P., Marchant, J. L., Hassabis, D., Weiskopf, N., Seymour, B., … Frith, C. D. (2007). When fear is near: threat imminence elicits prefrontal-periaqueductal gray shifts in humans. *Science*, *317*(5841), 1079–1083. doi: 10.1126/science.1144298

Mongillo, E. A., Irwin, J. R., Whalen, D. H., Klaiman, C., Carter, A. S., & Schultz, R. T. (2008). Audiovisual processing in children with and without autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *38*(7), 1349–1358. doi: 10.1007/s10803-007-0521-y

Mooney, C. M. (1957). Age in the development of closure ability in children. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *11*(4), 219–226. doi: 10.1037/h0083717

Moors, A. (2010). Automatic constructive appraisal as a candidate cause of emotion. *Emotion Review*, *2*(2), 139–156. doi: 10.1177/1754073909351755

Moran, R. J., Campo, P., Symmonds, M., Stephan, K. E., Dolan, R. J., & Friston, K. J. (2013). Free energy, precision and learning: The role of cholinergic neuromodulation. *The Journal of Neuroscience*, *33*(19), 8227–8236. doi: 10.1523/JNEUROSCI.4255-12.2013

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to cousineau (2005). *Tutorial in Quantitative Methods for Psychology*, *4*(2), 61–64.

Moriya, J., & Sugiura, Y. (2012). High visual working memory capacity in trait social anxiety. *PLoS ONE*, *7*(4), e34244. doi: 10.1371/journal.pone.0034244

Moriya, J., & Tanno, Y. (2009). Competition between endogenous and exogenous attention to nonemotional stimuli in social anxiety. *Emotion*, *9*(5), 739–743. doi: 10.1037/a0016817

Mottron, L., Bouvet, L., Bonnel, A., Samson, F., Burack, J. A., Dawson, M., & Heaton, P. (2013). Veridical mapping in the development of exceptional autistic abilities. *Neuroscience & Biobehavioral Reviews*, *37*(2), 209–228. doi: 10.1016/j.neubiorev.2012.11.016

Mottron, L., & Burack, J. A. (2001). Enhanced perceptual functioning in the development of autism. In J. A. Burack, T. Charman, N. Yirmiya, & P. R. Zelazo (Eds.), *The development of autism: Perspectives from theory and research* (pp. 131–148). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Mottron, L., Dawson, M., Soulières, I., Hubert, B., & Burack, J. (2006). Enhanced perceptual functioning in autism: An update, and eight principles of autistic perception. *Journal of Autism and Developmental Disorders*, *36*(1), 27–43.

Mottron, L., Peretz, I., & Menard, E. (2000). Local and global processing of music in high-functioning persons with autism: beyond central coherence? *Journal of Child Psychology & Psychiatry*, *41*(0021-9630 (Print)), 1057–1065.

Moutoussis, M., Fearon, P., El-Deredy, W., Dolan, R. J., & Friston, K. J. (2014). Bayesian inferences about the self (and others): A review. *Consciousness and Cognition*, *25*, 67–76. doi: 10.1016/j.concog.2014.01.009

Muckli, L., Petro, L. S., & Smith, F. W. (2013). Backwards is the way forward: feedback in the cortical hierarchy predicts the expected future. *The Behavioral and Brain Sciences*, *36*(3), 221. doi: 10.1017/S0140525X12002361

Murray, E. A. (2007). The amygdala, reward and emotion. *Trends in Cognitive Sciences*, *11*(11), 489–497. doi: 10.1016/j.tics.2007.08.013

Murray, S. O., Kersten, D., Olshausen, B. A., Schrater, P., & Woods,

D. L. (2002). Shape perception reduces activity in human primary visual cortex. *Proceedings of the National Academy of Sciences*, *99*(23), 15164–15169.

Mushtaq, F., Bland, A. R., & Schaefer, A. (2011). Uncertainty and cognitive control. *Frontiers in Psychology*, *2*. doi: 10.3389/fpsyg.2011.00249

Muth, C., & Carbon, C. (2013). The aesthetic aha: On the pleasure of having insights into gestalt. *Acta Psychologica*, *144*(1), 25–30. doi: 10.1016/j.actpsy.2013.05.001

Nackaerts, E., Wagemans, J., Helsen, W., Swinnen, S. P., Wenderoth, N., & Alaerts, K. (2012). Recognizing biological motion and emotions from Point-Light displays in autism spectrum disorders. *PLoS ONE*, *7*(9), e44473. doi: 10.1371/journal.pone.0044473

Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., & Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature Neuroscience*, *15*(7), 1040–1046. doi: 10.1038/nn.3130

Nemeth, D., Janacsek, K., Balogh, V., Londe, Z., Mingesz, R., Fazekas, M., ... Vetro, A. (2010). Learning in autism: implicitly superb. *PloS one*, *5*(7), e11731. doi: 10.1371/journal.pone.0011731

Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium.

Nieuwenhuis, S., Forstmann, B. U., & Wagenmakers, E. (2011). Erroneous analyses of interactions in neuroscience: a problem of significance. *Nature Neuroscience*, *14*(9), 1105–1107. doi: 10.1038/nn.2886

Nikolić, D. (2010). The brain is a context machine. *Review of Psychology*, *17*(1), 33–38.

Niv, Y., & Chan, S. (2011). On the value of information and other rewards. *Nature Neuroscience*, *14*(9), 1095–1097. doi: 10.1038/nn.2918

Noordewier, M. K., & Breugelmans, S. M. (2013). On the valence of surprise. *Cognition & Emotion*, *27*(7), 1326–1334. doi: 10.1080/02699931.2013.777660

Norton, M. I., Mochon, D., & Ariely, D. (2012). The IKEA effect: When labor leads to love. *Journal of Consumer Psychology*, *22*(3), 453–460. doi: 10.1016/j.jcps.2011.08.002

Noyce, A., & Sekuler, R. (2011). Surprises are mistakes: An EEG source localization study of prediction errors. *Journal of Vision*, *11*(11), 182. doi: 10.1167/11.11.182

O'Connor, K. (2012). Auditory processing in autism spectrum disorder: a review. *Neuroscience and Behavioral Reviews*, *36*(1873-7528

(Electronic)), 836–854. doi: 10.1016/j.neubiorev.2011.11.008

Odom, S. L., Brown, W. H., Frey, T., Karasu, N., Smith-Canter, L. L., & Strain, P. S. (2003). Evidence-Based practices for young children with autism contributions for Single-Subject design research. *Focus on Autism and Other Developmental Disabilities*, *18*(3), 166–175. doi: 10.1177/10883576030180030401

Ogawa, H., & Watanabe, K. (2011). Implicit learning increases preference for predictive visual display. *Attention, Perception, & Psychophysics*, *73*, 1815–1822. doi: 10.3758/s13414-010-0041-2

Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review*, *108*(3), 483–522. doi: 10.1037/0033-295X.108.3.483

O'Reilly, J. X. (2013). Making predictions in a changing world—inference, uncertainty, and learning. *Frontiers in Decision Neuroscience*, *7*, 105. doi: 10.3389/fnins.2013.00105

O'Riordan, M. A., Plaisted, K. C., Driver, J., & Baron-Cohen, S. (2001). Superior visual search in autism. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(3), 719–730. doi: 10.1037/0096-1523.27.3.719

Ornitz, D. E. M. (1974). The modulation of sensory input and motor output in autistic children. *Journal of autism and childhood schizophrenia*, *4*(3), 197–215. doi: 10.1007/BF02115226

Oudeyer, P., Baranes, A., & Kaplan, F. (2010). Intrinsically motivated exploration for developmental and active sensorimotor learning. In O. Sigaud & J. Peters (Eds.), *From motor learning to interaction learning in robots* (pp. 107–146). Springer Berlin Heidelberg.

Oudeyer, P., Kaplan, F., & Hafner, V. (2007). Intrinsic motivation systems for autonomous mental development. *IEEE Transactions on Evolutionary Computation*, *11*(2), 265–286. doi: 10.1109/TEVC.2006.890271

Ozonoff, S., Macari, S., Young, G. S., Goldring, S., Thompson, M., & Rogers, S. J. (2008). Atypical object exploration at 12 months of age is associated with autism in a prospective sample. *Autism*, *12*(5), 457–472. doi: 10.1177/1362361308096402

Palmer, C. J., Paton, B., Hohwy, J., & Enticott, P. G. (2013). Movement under uncertainty: The effects of the rubber-hand illusion vary along the nonclinical autism spectrum. *Neuropsychologia*, *51*(10), 1942–1951. doi: 10.1016/j.neuropsychologia.2013.06.020

Panksepp, J. (2005). Affective consciousness: Core emotional feelings in animals and humans. *Consciousness and Cognition*, *14*(1), 30–80. doi: 10.1016/j.concog.2004.10.004

Parrott, W. G., & Gleitman, H. (1989). Infants' expectations in play:

The joy of peek-a-boo. *Cognition & Emotion*, *3*(4), 291–311. doi: 10.1080/02699938908412710

Paton, B., Hohwy, J., & Enticott, P. G. (2012). The rubber hand illusion reveals proprioceptive and sensorimotor differences in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *42*(9), 1870–1883. doi: 10.1007/s10803-011-1430-7

Pavlova, M. A. (2012). Biological motion processing as a hallmark of social cognition. *Cerebral Cortex*, *22*(5), 981–995. doi: 10.1093/cercor/bhr156

Payzan-LeNestour, E., Dunne, S., Bossaerts, P., & O'Doherty, J. P. (2013). The neural representation of unexpected uncertainty during Value-Based decision making. *Neuron*, *79*(1), 191–201. doi: 10.1016/j.neuron.2013.04.037

Peirce, J. W. (2007). PsychoPy—Psychophysics software in python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. doi: 10.1016/j.jneumeth.2006.11.017

Pellicano, E. (2011). Psychological models of autism: An overview. In I. Roth & P. Rezaie (Eds.), *Researching the autism spectrum: Contemporary perspectives* (pp. n–n). Cambridge, UK: Cambridge University Press.

Pellicano, E., & Burr, D. (2012). When the world becomes 'too real': a Bayesian explanation of autistic perception. *Trends in Cognitive Sciences*, *16*(10), 504–510. doi: 10.1016/j.tics.2012.08.009

Pellicano, E., Gibson, L., Maybery, M., Durkin, K., & Badcock, D. R. (2005). Abnormal global processing along the dorsal visual pathway in autism: a possible mechanism for weak visuospatial coherence? *Neuropsychologia*, *43*(7), 1044–1053. doi: 10.1016/j.neuropsychologia.2004.10.003

Pellicano, E., Jeffery, L., Burr, D., & Rhodes, G. (2007). Abnormal adaptive face-coding mechanisms in children with autism spectrum disorder. *Current Biology*, *17*(0960-9822 (Print)), 1508–1512. doi: 10.1016/j.cub.2007.07.065

Pelowski, M., & Akiba, F. (2011). A model of art perception, evaluation and emotion in transformative aesthetic experience. *New Ideas in Psychology*, *29*(2), 80–97. doi: 10.1016/j.newideapsych.2010.04.001

Pepperell, R. (2011). Connecting art and the brain: An artist's perspective on visual indeterminacy. *Frontiers in Human Neuroscience*, *5*. doi: 10.3389/fnhum.2011.00084

Perry, E. K., Lee, M. L., Martin-Ruiz, C. M., Court, J. A., Volsen, S. G., Merrit, J., ... Wenk, G. L. (2001). Cholinergic activity in autism: Abnormalities in the cerebral cortex and basal forebrain. *American Journal of Psychiatry*, *158*(7), 1058–1066. doi: 10.1176/appi.ajp.158.7.1058

Peterson, M. A. (1994). Object recognition processes can and do operate before Figure-Ground organization. *Current Directions in Psychological Science*, *3*(4), 105–111. doi: 10.1111/j.1467-8721.1994.tb00156.x

Phaf, R. H., & Rotteveel, M. (2012). Affective monitoring: A generic mechanism for affect elicitation. *Frontiers in Psychology*, *3*. doi: 10.3389/fpsyg.2012.00047

Phillips, W. A. (2012). Self-Organized complexity and coherent infomax from the viewpoint of jaynes's probability theory. *Information*, *3*(1), 1–15. doi: 10.3390/info3010001

Phillips, W. A. (2013). Neuronal inference must be local, selective, and coordinated. *Behavioral and Brain Sciences*, *36*(03), 222–223. doi: 10.1017/S0140525X12002257

Piaget, J. (1937). *La construction du réel chez l'enfant. [The construction of the real in the child.]*. Oxford, England: Delachaux & Niestle.

Picard, F. (2013). State of belief, subjective certainty and bliss as a product of cortical dysfunction. *Cortex*, *49*(9), 2494–2500. doi: 10.1016/j.cortex.2013.01.006

Pierce, K., Glad, K. S., & Schreibman, L. (1997). Social perception in children with autism: An attentional deficit? *Journal of Autism and Developmental Disorders*, *27*(3), 265–282. doi: 10.1023/A:1025898314332

Pinborough-Zimmerman, J., Bakian, A. V., Fombonne, E., Bilder, D., Taylor, J., & McMahon, W. M. (2012). Changes in the administrative prevalence of autism spectrum disorders: Contribution of special education and health from 2002-2008. *Journal of Autism and Developmental Disorders*, *42*(4), 521–530. doi: 10.1007/s10803-011-1265-2

Pinker, S. (2003). *The blank slate: The modern denial of human nature*. Penguin Books.

Pinna, B. (2010). New gestalt principles of perceptual organization: an extension from grouping to shape and meaning. *Gestalt Theory*, *32*, 1–67.

Plaisted, K. C. (2001). Reduced generalization in autism: An alternative to weak central coherence. In J. A. Burack, T. Charman, N. Yirmiya, & P. R. Zelazo (Eds.), *The development of autism: Perspectives from theory and research* (pp. 149–169). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Plaisted, K. C., O'Riordan, M., & Baron-Cohen, S. (1998a). Enhanced discrimination of novel, highly similar stimuli by adults with autism during a perceptual learning task. *Journal of Child Psychology and Psychiatry*, *39*(5), 765–775. doi: 10.1111/1469-7610.00375

Plaisted, K. C., O'Riordan, M., & Baron-Cohen, S. (1998b). Enhanced

visual search for a conjunctive target in autism: A research note. *Journal of Child Psychology and Psychiatry*, *39*(5), 777–783. doi: 10.1111/1469-7610.00376

Polani, D. (2009). Information: Currency of life? *HFSP Journal*, *3*(5), 307–316. doi: 10.2976/1.3171566

Poljac, E., & Bekkering, H. (2012). A review of intentional and cognitive control in autism. *Frontiers in Psychology*, *3:436*. doi: 10.3389/fpsyg.2012.00436

Powers, W. T. (1973). *Behavior: The control of perception* (Vol. xi). Oxford, England: Aldine.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(04), 515–526. doi: 10.1017/S0140525X00076512

Preuschoff, K., & Bossaerts, P. (2007). Adding prediction risk to the theory of reward learning. *Annals of the New York Academy of Sciences*, *1104*, 135–146. doi: 10.1196/annals.1390.005

Preuschoff, K., Quartz, S. R., & Bossaerts, P. (2008). Human insula activation reflects risk prediction errors as well as risk. *The Journal of Neuroscience*, *28*(11), 2745 –2752. doi: 10.1523/JNEUROSCI.4286-07.2008

Proulx, T., & Heine, S. J. (2009). Connections from kafka exposure to meaning threats improves implicit learning of an artificial grammar. *Psychological Science*, *20*(9), 1125–1131. doi: 10.1111/j.1467-9280.2009.02414.x

Proulx, T., Heine, S. J., & Vohs, K. D. (2010). When is the unfamiliar the uncanny? meaning affirmation after exposure to absurdist literature, humor, and art. *Personality and Social Psychology Bulletin*, *36*(6), 817 –829. doi: 10.1177/0146167210369896

Proulx, T., Inzlicht, M., & Harmon-Jones, E. (2012). Understanding all inconsistency compensation as a palliative response to violated expectations. *Trends in Cognitive Sciences*, *16*(5), 285–291. doi: 10.1016/j.tics.2012.04.002

Pruett, J., John R, LaMacchia, A., Hoertel, S., Squire, E., McVey, K., Todd, R. D., ... Petersen, S. E. (2011). Social and non-social cueing of visuospatial attention in autism and typical development. *Journal of Autism and Developmental Disorders*, *41*(6), 715–731. doi: 10.1007/s10803-010-1090-z

Purves, D., Wojtach, W. T., & Lotto, R. B. (2011). Understanding vision in wholly empirical terms. *Proceedings of the National Academy of Sciences*, *108*(Supplement_3), 15588–15595. doi: 10.1073/pnas.1012178108

R Core Team. (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rajkowski, J., Kubiak, P., & Aston-Jones, G. (1993). Correlations between locus coeruleus (LC) neural activity, pupil diameter and behavior in monkey support a role of LC in attention. In *Society for neuroscience abstracts* (Vol. 19, p. 974).

Ramachandran, V. S., & Hirstein, W. (1999). The science of art: a neurological theory of aesthetic experience. *Journal of Consciousness Studies*, *6*(6-7), 15–51.

Rao, R. P. N. (1999). An optimal estimation approach to visual perception and learning. *Vision Research*, *39*(11), 1963–1989. doi: 10.1016/S0042-6989(98)00279-X

Rao, R. P. N., & Ballard, D. H. (1997). Dynamic model of visual recognition predicts neural response properties in the visual cortex. *Neural Computation*, *9*(4), 721–763. doi: 10.1162/neco.1997.9.4.721

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. doi: 10.1038/4580

Rasmussen, D., & Eliasmith, C. (2013). God, the devil, and the details: Fleshing out the predictive processing framework. *Behavioral and Brain Sciences*, *36*(03), 223–224. doi: 10.1017/S0140525X12002154

Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: is beauty in the perceiver's processing experience? *Personality and Social Psychology Review*, *8*(4), 364–82. doi: 10.1207/s15327957pspr0804_3

Reber, R., Winkielman, P., & Schwarz, N. (1998). Effects of perceptual fluency on affective judgments. *Psychological Science*, *9*(1), 45–48. doi: 10.1111/1467-9280.00008

Redies, C. (2007). A universal model of esthetic perception based on the sensory coding of natural stimuli. *Spatial Vision*, *21*, 97–117. doi: 10.1163/156856807782753886

Reinberg, A., & Smolensky, M. H. (1983). *Introduction to chronobiology*. Springer.

Reisenzein, R. (1983). The schachter theory of emotion: Two decades later. *Psychological Bulletin*, *94*(2), 239–264. doi: 10.1037/0033-2909.94.2.239

Reisenzein, R. (2009). Emotional experience in the computational Belief–Desire theory of emotion. *Emotion Review*, *1*(3), 214–222. doi: 10.1177/1754073909103589

Reynhout, G., & Carter, M. (2011). Evaluation of the efficacy of social stories (TM) using three single subject metrics. *Research in Autism Spectrum Disorders*, *5*(2), 885–900. doi: 10.1016/j.rasd.2010.10.003

Rice, K., Moriuchi, J. M., Jones, W., & Klin, A. (2012). Parsing heterogeneity in autism spectrum disorders: Visual scanning of dynamic social scenes in school-aged children. *Journal of the American Academy of Child and Adolescent Psychiatry*, *51*(3), 238–248. doi: 10.1016/j.jaac.2011.12.017

Richler, J., Huerta, M., Bishop, S. L., & Lord, C. (2010). Developmental trajectories of restricted and repetitive behaviors and interests in children with autism spectrum disorders. *Development and Psychopathology*, *22*(01), 55. doi: 10.1017/S0954579409990265

Robertson, C. E., Kravitz, D. J., Freyberg, J., Baron-Cohen, S., & Baker, C. I. (2013). Slower rate of binocular rivalry in autism. *The Journal of Neuroscience*, *33*(43), 16983–16991. doi: 10.1523/JNEUROSCI.0448-13.2013

Robertson, C. E., Martin, A., Baker, C. I., & Baron-Cohen, S. (2012). Atypical integration of motion signals in autism spectrum conditions. *PLoS ONE*, *7*(11), e48173. doi: 10.1371/journal.pone.0048173

Robinson, M. J. F., Anselme, P., Fischer, A. M., & Berridge, K. C. (2014). Initial uncertainty in pavlovian reward prediction persistently elevates incentive salience and extends sign-tracking to normally unattractive cues. *Behavioural Brain Research*, *266*, 119–130. doi: 10.1016/j.bbr.2014.03.004

Rogers, S. J., & Vismara, L. A. (2008). Evidence-Based comprehensive treatments for early autism. *Journal of clinical child and adolescent psychology*, *37*(1), 8–38. doi: 10.1080/15374410701817808

Rommelse, N. N., Geurts, H. M., Franke, B., Buitelaar, J. K., & Hartman, C. A. (2011). A review on cognitive and brain endophenotypes that may be common in autism spectrum disorder and attention-deficit/hyperactivity disorder and facilitate the search for pleiotropic genes. *Neuroscience & Biobehavioral Reviews*, *35*(6), 1363–1396. doi: 10.1016/j.neubiorev.2011.02.015

Ropar, D., & Mitchell, P. (2001). Susceptibility to illusions and performance on visuospatial tasks in individuals with autism. *Journal of Child Psychology and Psychiatry*, *42*(4), 539–549. doi: 10.1111/1469-7610.00748

Ropar, D., & Mitchell, P. (2002). Shape constancy in autism: the role of prior knowledge and perspective cues. *Journal of Child Psychology and Psychiatry*, *43*(5), 647–653. doi: 10.1111/1469-7610.00053

Ropar, D., Mitchell, P., & Ackroyd, K. (2003). Do children with autism find it difficult to offer alternative interpretations to ambiguous figures? *British Journal of Developmental Psychology*,

21(3), 387–395. doi: 10.1348/026151003322277766

Ropar, D., & Peebles, D. (2007). Sorting preference in children with autism: The dominance of concrete features. *Journal of Autism and Developmental Disorders*, *37*(2), 270–280. doi: 10.1007/s10803-006-0166-2

Rosen, R. (2012). *Anticipatory systems: Philosophical, mathematical, and methodological foundations*. Springer Science & Business Media.

Rouse, H., Donnelly, N., Hadwin, J. A., & Brown, T. (2004). Do children with autism perceive second-order relational features? the case of the thatcher illusion. *Journal of Child Psychology and Psychiatry*, *45*(7), 1246–1257. doi: 10.1111/j.1469-7610.2004.00317.x

Rozin, P., & Kennel, K. (1983). Acquired preferences for piquant foods by chimpanzees. *Appetite*, *4*(2), 69–77. doi: 10.1016/S0195-6663(83)80003-8

Rozin, P., Rozin, A., Appel, B., & Wachtel, C. (2006). Documenting and explaining the common AAB pattern in music and humor: Establishing and breaking expectations. *Emotion*, *6*(3), 349–355. doi: 10.1037/1528-3542.6.3.349

Rushworth, M., Behrens, T., Rudebeck, P., & Walton, M. (2007). Contrasting roles for cingulate and orbitofrontal cortex in decisions and social behaviour. *Trends in Cognitive Sciences*, *11*(4), 168–176. doi: 10.1016/j.tics.2007.01.004

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145–172. doi: 10.1037/0033-295X.110.1.145

Rutherford, M. D., Troubridge, E. K., & Walsh, J. (2012). Visual afterimages of emotional faces in high functioning autism. *Journal of Autism and Developmental Disorders*, *42*(2), 221–229. doi: 10.1007/s10803-011-1233-x

Said, C. P., Dotsch, R., & Todorov, A. (2010). The amygdala and FFA track both social and non-social face dimensions. *Neuropsychologia*, *48*(12), 3596–3605. doi: 10.1016/j.neuropsychologia.2010.08.009

Said, C. P., Egan, R. D., Minshew, N. J., Behrmann, M., & Heeger, D. J. (2013). Normal binocular rivalry in autism: Implications for the excitation/inhibition imbalance hypothesis. *Vision Research*, *77*, 59–66. doi: 10.1016/j.visres.2012.11.002

Sander, D., Grafman, J., & Zalla, T. (2003). The human amygdala: an evolved system for relevance detection. *Reviews in the neurosciences*, *14*(4).

Sarinopoulos, I., Grupe, D. W., Mackiewicz, K. L., Herrington, J. D., Lor, M., Steege, E. E., & Nitschke, J. B. (2010). Uncertainty during anticipation modulates neural responses to aversion in

human insula and amygdala. *Cerebral Cortex*, *20*(4), 929–940. doi: 10.1093/cercor/bhp155

Sasson, N. J., Elison, J. T., Turner-Brown, L. M., Dichter, G. S., & Bodfish, J. W. (2011). Brief report: Circumscribed attention in young children with autism. *Journal of Autism and Developmental Disorders*, *41*(2), 242–247. doi: 10.1007/s10803-010-1038-3

Sasson, N. J., Turner-Brown, L. M., Holtzclaw, T. N., Lam, K. S., & Bodfish, J. W. (2008). Children with autism demonstrate circumscribed attention during passive viewing of complex social and nonsocial picture arrays. *Autism Research*, *1*(1), 31–42. doi: 10.1002/aur.4

Saygin, A., Cook, J., & Blakemore, S. (2010). Unaffected perceptual thresholds for biological and non-biological form-from-motion perception in autism spectrum conditions. *PLOS ONE*, *5*(10), e13491. doi: 10.1371/journal.pone.0013491

Schachter, S., & Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review*, *69*(5), 379–399. doi: 10.1037/h0046234

Schafer, R., & Murphy, G. (1943). The role of autism in a visual figure-ground relationship. *Journal of Experimental Psychology*, *32*(4), 335–343. doi: 10.1037/h0057755

Scherer, K. R., Schorr, A., & Johnstone, T. (2001). *Appraisal processes in emotion: Theory, methods, research (Series in affective science)*. Oxford University Press, USA.

Schiff, W., Caviness, J. A., & Gibson, J. J. (1962). Persistent fear responses in rhesus monkeys to the optical stimulus of "looming.". *Science*, *136*, 982–983. doi: 10.1126/science.136.3520.982

Schiffer, A., Ahlheim, C., Wurm, M. F., & Schubotz, R. I. (2012). Surprised at all the entropy: Hippocampal, caudate and midbrain contributions to learning from prediction errors. *PLoS ONE*, *7*(5), e36445. doi: 10.1371/journal.pone.0036445

Schiffer, A., & Schubotz, R. I. (2011). Caudate nucleus signals for breaches of expectation in a movement observation paradigm. *Frontiers in Human Neuroscience*, *5*. doi: 10.3389/fnhum.2011.00038

Schmack, K., Castro, A. G. d., Rothkirch, M., Sekutowicz, M., Rössler, H., Haynes, J., ... Sterzer, P. (2013). Delusions and the role of beliefs in perceptual inference. *The Journal of Neuroscience*, *33*(34), 13701–13712. doi: 10.1523/JNEUROSCI.1778-13.2013

Schmidhuber, J. (2010). Formal theory of creativity, fun, and intrinsic motivation (1990–2010). *Autonomous Mental Development, IEEE Transactions on*, *2*(3), 230–247.

Scholl, B. J. (2005). Innateness and (Bayesian) visual perception. *The*

innate mind: Structure and contents, 34.

Schouppe, N., Braem, S., Houwer, J. D., Silvetti, M., Verguts, T., Ridderinkhof, K. R., & Notebaert, W. (2014). No pain, no gain: the affective valence of congruency conditions changes following a successful response. *Cognitive, Affective, & Behavioral Neuroscience*, 1–11. doi: 10.3758/s13415-014-0318-3

Schouten, B., Troje, N. F., & Verfaillie, K. (2011). The facing bias in biological motion perception: structure, kinematics, and body parts. *Attention, Perception, & Psychophysics*, *73*(1), 130–143. doi: 10.3758/s13414-010-0018-1

Schouten, B., & Verfaillie, K. (2010). Determining the point of subjective ambiguity of ambiguous biological-motion figures with perspective cues. *Behavior Research Methods*, *42*(1), 161–167. doi: 10.3758/BRM.42.1.161

Schultz, W. (2007). Reward. *Scholarpedia*, *2*(3), 1652. doi: 10.4249/scholarpedia.1652

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593–1599. doi: 10.1126/science.275.5306.1593

Schulz, R. (1976). Effects of control and predictability on the physical and psychological well-being of the institutionalized aged. *Journal of Personality and Social Psychology*, *33*(5), 563–573. doi: 10.1037/0022-3514.33.5.563

Schwartenbeck, P., FitzGerald, T. H. B., Mathys, C., Dolan, R., & Friston, K. (2014). The dopaminergic midbrain encodes the expected certainty about desired outcomes. *Cerebral Cortex*, bhu159. doi: 10.1093/cercor/bhu159

Sengupta, B., Stemmler, M. B., & Friston, K. J. (2013). Information and efficiency in the nervous System—A synthesis. *PLoS Computional Biology*, *9*(7), e1003157. doi: 10.1371/journal.pcbi.1003157

Serences, J. T. (2008). Value-based modulations in human visual cortex. *Neuron*, *60*(6), 1169–81. doi: 10.1016/j.neuron.2008.10.051

Serretti, A., & Fabbri, C. (2013). Shared genetics among major psychiatric disorders. *Lancet*, *381*(9875), 1339–1341. doi: 10.1016/S0140-6736(13)60223-8

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, *17*(11), 565–573. doi: 10.1016/j.tics.2013.09.007

Seth, A. K., Suzuki, K., & Critchley, H. D. (2012). An interoceptive predictive coding model of conscious presence. *Frontiers in Psychology*, *2*. doi: 10.3389/fpsyg.2011.00395

Shackman, A. J., Salomons, T. V., Slagter, H. A., Fox, A. S., Winter,

J. J., & Davidson, R. J. (2011). The integration of negative affect, pain and cognitive control in the cingulate cortex. *Nature Reviews Neuroscience*, *12*(3), 154–167. doi: 10.1038/nrn2994

Shah, A., & Frith, U. (1993). Why do autistic individuals show superior performance on the block design task? *Journal of Child Psychology and Psychiatry*, *34*(8), 1351–1364. doi: 10.1111/j.1469-7610.1993.tb02095.x

Silverstein, S. M. (2013). Schizophrenia-related phenomena that challenge prediction error as the basis of cognitive functioning. *Behavioral and Brain Sciences*, *36*(03), 229–230. doi: 10.1017/S0140525X12002221

Silvia, P. J. (2001). Interest and interests: The psychology of constructive capriciousness. *Review of General Psychology*, *5*(3), 270–290. doi: 10.1037/1089-2680.5.3.270

Silvia, P. J. (2008). Interest—The curious emotion. *Current Directions in Psychological Science*, *17*(1), 57–60. doi: 10.1111/j.1467-8721.2008.00548.x

Simmons, D. R., Robertson, A. E., McKay, L. S., Toal, E., McAleer, P., & Pollick, F. E. (2009). Vision in autism spectrum disorders. *Vision Research*, *49*(22), 2705–2739. doi: 10.1016/j.visres.2009.08.005

Singer, T., Critchley, H. D., & Preuschoff, K. (2009). A common role of insula in feelings, empathy and uncertainty. *Trends in Cognitive Sciences*, *13*(8), 334–340. doi: 10.1016/j.tics.2009.05.001

Singh, S., Lewis, R. L., & Barto, A. G. (2009). Where do rewards come from. In *Proceedings of the annual conference of the cognitive science society* (p. 2601–2606).

Sloboda, J. A. (1991). Music structure and emotional response: Some empirical findings. *Psychology of Music*, *19*(2), 110 –120. doi: 10.1177/0305735691192002

Smith, E. G., & Bennetto, L. (2007). Audiovisual speech integration and lipreading in autism. *Journal of Child Psychology and Psychiatry*, *48*(8), 813–821. doi: 10.1111/j.1469-7610.2007.01766.x

Smith, J. D. (2009). The study of animal metacognition. *Trends in Cognitive Sciences*, *13*(9), 389–396. doi: 10.1016/j.tics.2009.06.009

Smith, J. D., Coutinho, M. V. C., Church, B. A., & Beran, M. J. (2013). Executive-attentional uncertainty responses by rhesus macaques (Macaca mulatta). *Journal of Experimental Psychology: General*, *142*(2), 458–475. doi: 10.1037/a0029601

Solway, A., & Botvinick, M. M. (2012). Goal-directed decision making as probabilistic inference: A computational framework and potential neural correlates. *Psychological Review*, *119*(1), 120–154. doi: 10.1037/a0026435

Soulieres, I., Mottron, L., Giguere, G., & Larochelle, S. (2011). Category induction in autism: slower, perhaps different, but certainly possible. *Quarterly Journal of Experimental Psychology*, *64*(1747-0226 (Electronic)), 311–327. doi: 10.1080/17470218.2010.492994

Spencer, J., O'Brien, J., Riggs, K., Braddick, O., Atkinson, J., & Wattam-Bell, J. (2000). Motion processing in autism: Evidence for a dorsal stream deficiency. *Neuroreport*, *11*(12), 2765–2767.

Spengler, S., Bird, G., & Brass, M. (2010). Hyperimitation of actions is related to reduced understanding of others' minds in autism spectrum conditions. *Biological Psychiatry*, *68*(12), 1148–1155. doi: 10.1016/j.biopsych.2010.09.017

Sperber, D. (2005). Modularity and relevance: How can a massively modular mind be flexible and context-sensitive? In P. Carruthers & S. Laurence (Eds.), *The innate mind: Structure and contents.* USA: Oxford University Press.

Sperber, D., & Hirschfeld, L. A. (2004). The cognitive foundations of cultural stability and diversity. *Trends in Cognitive Sciences*, *8*(1), 40–46. doi: 10.1016/j.tics.2003.11.002

Spinoza, B. d. (2001). *Ethics*. Wordsworth Editions.

Spratling, M. W. (2010). Predictive coding as a model of response properties in cortical area v1. *The Journal of Neuroscience*, *30*(9), 3531–3543. doi: 10.1523/JNEUROSCI.4911-09.2010

Spruyt, A., De Houwer, J., Hermans, D., & Eelen, P. (2007). Affective priming of nonaffective semantic categorization responses. *Experimental Psychology*, *54*(1), 44–53. doi: 10.1027/1618-3169.54.1.44

Sroufe, L. A., & Waters, E. (1976). The ontogenesis of smiling and laughter: A perspective on the organization of development in infancy. *Psychological Review*, *83*(3), 173–189. doi: 10.1037/0033-295X.83.3.173

Stanisor, L., van der Togt, C., Pennartz, C. M. A., & Roelfsema, P. R. (2013). A unified selection signal for attention and reward in primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(22), 9136–9141. doi: 10.1073/pnas.1300117110

Stansbury, D. E., Naselaris, T., & Gallant, J. L. (2013). Natural scene statistics account for the representation of scene categories in human visual cortex. *Neuron*, *79*(5), 1025–1034. doi: 10.1016/j.neuron.2013.06.034

Stefanucci, J. K., Proffitt, D. R., Clore, G. L., & Parekh, N. (2008). Skating down a steeper slope: Fear influences the perception of geographical slant. *Perception*, *37*(2), 321–323. doi: 10.1068/p5796

Stephan, K. E., Friston, K. J., & Frith, C. D. (2009). Dysconnection in schizophrenia: From abnormal synaptic plasticity to failures of self-monitoring. *Schizophrenia Bulletin*, *35*(3), 509–527. doi: 10.1093/schbul/sbn176

Sterling, P. (2012). Allostasis: A model of predictive regulation. *Physiology & Behavior*, *106*(1), 5–15. doi: 10.1016/j.physbeh.2011.06.004

Stewart, L. H., Ajina, S., Getov, S., Bahrami, B., Todorov, A., & Rees, G. (2012). Unconscious evaluation of faces on social dimensions. *Journal of Experimental Psychology: General*, *141*(4), 715–727. doi: 10.1037/a0027950

Stoet, G., & López, B. (2011). Task-switching abilities in children with autism spectrum disorder. *European Journal of Developmental Psychology*, *8*(2), 244–260. doi: 10.1080/17405629.2010.492000

Stolarova, M., Keil, A., & Moratti, S. (2006). Modulation of the c1 visual event-related component by conditioned stimuli: evidence for sensory plasticity in early affective perception. *Cerebral Cortex*, *16*(6), 876–87. doi: 10.1093/cercor/bhj031

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience*, *11*(9), 1004–1006. doi: 10.1038/nn.2163

Sutton, R. S., & Barto, A. G. (1981). Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review*, *88*(2), 135–170. doi: 10.1037/0033-295X.88.2.135

Synofzik, M., Lindner, A., & Thier, P. (2008). The cerebellum updates predictions about the visual consequences of one's behavior. *Current Biology*, *18*(11), 814–818. doi: 10.1016/j.cub.2008.04.071

Taatgen, N., & Anderson, J. R. (2010). The past, present, and future of cognitive architectures. *Topics in Cognitive Science*, *2*(4), 693–704. doi: 10.1111/j.1756-8765.2009.01063.x

Taylor, N., Isaac, C., & Milne, E. (2010). A comparison of the development of audiovisual integration in children with autism spectrum disorders and typically developing children. *Journal of Autism and Developmental Disorders*, *40*(11), 1403–1411. doi: 10.1007/s10803-010-1000-4

Teufel, C., Subramaniam, N., & Fletcher, P. C. (2013). The role of priors in Bayesian models of perception. *Frontiers in Computational Neuroscience*, 25. doi: 10.3389/fncom.2013.00025

Thompson, M. M., Naccarato, M. E., Parker, K. C., & Moskowitz, G. B. (2001). The personal need for structure and personal fear of invalidity measures: Historical perspectives, current ap-

plications, and future directions. In *Cognitive social psychology: The princeton symposium on the legacy and future of social cognition* (p. 19–39).

Thornton, C. (2014a). *Experiments in predictive coding* [Unpublished manuscript]. University of Sussex, Brighton, UK.

Thornton, C. (2014b). *Experiments in sparse-coded predictive processing* [Unpublished manuscript]. University of Sussex, Brighton, UK.

Tinbergen, N. (1951). *The study of instinct*. London, U.K.: Oxford Clarendon Press.

Todorovic, A., & de Lange, F. (2012). Repetition suppression and expectation suppression are dissociable in time in early auditory evoked fields. *Journal of Neuroscience*, *32*(1529-2401 (Electronic)), 13389–13395. doi: 10.1523/JNEUROSCI.2227-12.2012

Todorovic, A., van Ede, F., Maris, E., & de Lange, F. (2011). Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: an MEG study. *Journal of Neuroscience*, *31*(1529-2401 (Electronic)), 9118–9123. doi: 10.1523/JNEUROSCI.1425-11.2011

Tomchek, S. D., & Dunn, W. (2007). Sensory processing in children with and without autism: A comparative study using the short sensory profile. *American Journal of Occupational Therapy*, *61*(2), 190–200. doi: 10.5014/ajot.61.2.190

Topolinski, S. (2010). Moving the eye of the beholder. *Psychological Science*, *21*(9), 1220 –1224. doi: 10.1177/0956797610378308

Topolinski, S., & Reber, R. (2010). Gaining insight into the "Aha" experience. *Current Directions in Psychological Science*, *19*(6), 402–405. doi: 10.1177/0963721410388803

Treffert, D. A. (2009). The savant syndrome: an extraordinary condition. a synopsis: past, present, future. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1522), 1351–1357. doi: 10.1098/rstb.2008.0326

Troje, N. F. (2002). Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of Vision*, *2*(5), 371–387. doi: 10.1167/2.5.2

Turiano, N. A., Chapman, B. P., Agrigoroaei, S., Infurna, F. J., & Lachman, M. (2014). Perceived control reduces mortality risk at low, not high, education levels. *Health Psychology*, *33*(8), 883–890. doi: 10.1037/hea0000022

Turner, M. (1999). Annotation: Repetitive behaviour in autism: A review of psychological research. *Journal of Child Psychology and Psychiatry*, *40*(6), 839–849. doi: 10.1111/1469-7610.00502

Uddin, L. Q., & Menon, V. (2009). The anterior insula in autism: Under-connected and under-examined. *Neuroscience & Biobe-*

havioral Reviews, *33*(8), 1198–1203. doi: 10.1016/j.neubiorev.2009.06.002

Uljarevic, M., & Hamilton, A. (2013). Recognition of emotions in autism: A formal Meta-Analysis. *Journal of Autism and Developmental Disorders*, *43*(7), 1517–1526. doi: 10.1007/s10803-012-1695-5

Valdesolo, P., & Graham, J. (2014). Awe, uncertainty, and agency detection. *Psychological Science*, *25*(1), 170–178. doi: 10.1177/0956797613501884

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology*, *43*(2), 161–204.

Van de Cruys, S., de-Wit, L., Evers, K., Boets, B., & Wagemans, J. (2013). Weak priors versus overfitting of predictions in autism: Reply to pellicano and burr (TICS, 2012). *i-Perception*, *4*(2), 95–97.

van Boxtel, J. J. A., & Lu, H. (2012). Signature movements lead to efficient search for threatening actions. *PloS one*, *7*(5), e37085.

van Boxtel, J. J. A., & Lu, H. (2013). A predictive coding perspective on autism spectrum disorders. *Frontiers in Psychology*, *4*. doi: 10.3389/fpsyg.2013.00019

Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: Predictive coding in autism. *Psychological Review*(In press).

Van de Cruys, S., & Wagemans, J. (2011a). Gestalts as Predictions—Some reflections and an application to art. *Gestalt Theory*, *33*(3), 325.

Van de Cruys, S., & Wagemans, J. (2011b). Putting reward in art: A tentative prediction error account of visual art. *i-Perception*, *2*(9), 1035–1062. doi: 10.1068/i0466aap

Vandenbroucke, M., Scholte, H., van Engeland, H., Lamme, V., & Kemner, C. (2008). Coherent versus component motion perception in autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *38*(0162-3257 (Print)), 941–949. doi: 10.1007/s10803-007-0467-0

van der Geest, J. N., Kemner, C., Camfferman, G., Verbaten, M. N., & van Engeland, H. (2002). Looking at images with human figures: Comparison between autistic and normal children. *Journal of Autism and Developmental Disorders*, *32*(2), 69–75. doi: 10.1023/A:1014832420206

Van der Hallen, R., Evers, K., Brewaeys, K., Van den Noortgate, W., & Wagemans, J. (2014). Global processing takes time: A Meta-

Analysis on Local-Global visual processing in ASD. *Psychological Bulletin*.

van der Helm, P. A. (2000). Simplicity versus likelihood in visual perception: From surprisals to precisals. *Psychological Bulletin*, *126*(5), 770–800. doi: 10.1037/0033-2909.126.5.770

Vaneechoutte, M. (1993). The memetic basis of religion. *Nature*, *365*(6444), 290–290. doi: 10.1038/365290a0

Van Eylen, L., Boets, B., Steyaert, J., Evers, K., Wagemans, J., & Noens, I. (2011). Cognitive flexibility in autism spectrum disorder: Explaining the inconsistencies? *Research in Autism Spectrum Disorders*, *5*(4), 1390–1401. doi: 10.1016/j.rasd.2011.01.025

van Harreveld, F., Rutjens, B. T., Schneider, I. K., Nohlen, H. U., & Keskinis, K. (2014). In doubt and disorderly: Ambivalence promotes compensatory perceptions of order. *Journal of Experimental Psychology. General*, *143*(4), 1666–1676. doi: 10.1037/a0036099

van Leeuwen, C. (2007). What needs to emerge to make you conscious? *Journal of Consciousness Studies*, *14*(1-1), 115–136.

Vanrie, J., Dekeyser, M., & Verfaillie, K. (2004). Bistability and biasing effects in the perception of ambiguous point-light walkers. *Perception*, *33*(5), 547–560. doi: 10.1068/p5004

Varela, F. G., Maturana, H. R., & Uribe, R. (1974). Autopoiesis: The organization of living systems, its characterization and a model. *Biosystems*, *5*(4), 187–196. doi: 10.1016/0303-2647(74)90031-8

Verduzco-Flores, S., & O'Reilly, R. C. (2014). The cerebellum as a predictor of error increase in a controller. *arXiv:1408.3444 [q-bio]*.

Verstijnen, I., van Leeuwen, C., Goldschmidt, G., Hamel, R., & Hennessey, J. (1998). Sketching and creative discovery. *Design Studies*, *19*(4), 519–546. doi: 10.1016/S0142-694X(98)00017-9

Vessel, E. A. (2004). *Behavioral and neural investigation of perceptual affect* Unpublished doctoral dissertation. University of Southern California.

Vladusich, T., Olu-Lafe, O., Kim, D., Tager-Flusberg, H., & Grossberg, S. (2010). Prototypical category learning in high-functioning autism. *Autism Research*, *3*(5), 226–236. doi: 10.1002/aur.148

Volkmar, F., Chawarska, K., & Klin, A. (2005). Autism in infancy and early childhood. *Annual Review of Psychology*, *56*(1), 315–336. doi: 10.1146/annurev.psych.56.091103.070159

von Glasersfeld, E. (1991). An exposition of constructivism: Why some like it radical. In *Facets of systems science* (pp. 229–238).

Springer US.

von Glasersfeld, E. (1995). *Radical constructivism: A way of knowing and learning. studies in mathematics education series: 6.* Bristol, PA: Falmer Press, Taylor & Francis Inc.

Voss, A., Rothermund, K., & Brandtstädter, J. (2008). Interpreting ambiguous stimuli: Separating perceptual and judgmental biases. *Journal of Experimental Social Psychology*, *44*(4), 1048–1056. doi: 10.1016/j.jesp.2007.10.009

Vossel, S., Mathys, C., Daunizeau, J., Bauer, M., Driver, J., Friston, K. J., & Stephan, K. E. (2014). Spatial attention, precision, and Bayesian inference: a study of saccadic response speed. *Cerebral Cortex*, *24*(6), 1436–1450. doi: 10.1093/cercor/bhs418

Vuilleumier, P. (2005). How brains beware: neural mechanisms of emotional attention. *Trends in Cognitive Sciences*, *9*(12), 585–594. doi: 10.1016/j.tics.2005.10.011

Vygotsky, L. S. (1971). *Psychology of art*. MIT Press.

Wacongne, C., Changeux, J., & Dehaene, S. (2012). A neuronal model of predictive coding accounting for the mismatch negativity. *Journal of Neuroscience*, *32*(1529-2401 (Electronic)), 3665–3678. doi: 10.1523/JNEUROSCI.5003-11.2012

Wacongne, C., Labyt, E., Wassenhove, V. v., Bekinschtein, T., Naccache, L., & Dehaene, S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proceedings of the National Academy of Sciences*, *108*(51), 20754–20759. doi: 10.1073/pnas.1117807108

Wagemans, J., Elder, J., Kubovy, M., Palmer, S., Peterson, M., Singh, M., & von der, H. R. (2012). A century of gestalt psychology in visual perception: I. perceptual grouping and figure-ground organization 6. *Psychological Bulletin*, *138*(1939-1455 (Electronic)), 1172–1217. doi: 10.1037/a0029333

Wagemans, J., Feldman, J., Gepshtein, S., Kimchi, R., Pomerantz, J., van der Helm, P., & van, L. C. (2012). A century of gestalt psychology in visual perception: II. conceptual and theoretical foundations 7. *Psychological Bulletin*, *138*(1939-1455 (Electronic)), 1218–1252. doi: 10.1037/a0029334

Walther, D. B., Chai, B., Caddigan, E., Beck, D. M., & Fei-Fei, L. (2011). Simple line drawings suffice for functional MRI decoding of natural scene categories. *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.1015666108

Webb, S. J., Merkle, K., Murias, M., Richards, T., Aylward, E., & Dawson, G. (2012). ERP responses differentiate inverted but not upright face processing in adults with ASD. *Social Cognitive and Affective Neuroscience*, *7*(5), 578–587. doi: 10.1093/scan/nsp002

Weber, A., & Varela, F. J. (2002). Life after kant: Natural purposes and the autopoietic foundations of biological individuality. *Phenomenology and the Cognitive Sciences*, *1*(2), 97–125. doi: 10.1023/A:1020368120174

Weigelt, S., Koldewyn, K., & Kanwisher, N. (2012). Face identity recognition in autism spectrum disorders: a review of behavioral studies. *Neuroscience & Biobehavioral Reviews*, *36*(1873-7528 (Electronic)), 1060–1084. doi: 10.1016/j.neubiorev.2011.12.008

Wertheimer, M. (1923). Untersuchungen zur lehre der gestalt, II. *Psychologische Forschung*, *4*, 301–350.

Whalen, P. J. (1998). Fear, vigilance, and ambiguity: Initial neuroimaging studies of the human amygdala. *Current Directions in Psychological Science*, *7*(6), 177–188. doi: 10.1111/1467-8721.ep10836912

Whalen, P. J. (2007). The uncertainty of it all. *Trends in Cognitive Sciences*, *11*(12), 499–500. doi: 10.1016/j.tics.2007.08.016

White, S. J. (2013). The triple i hypothesis: Taking another('s) perspective on executive dysfunction in autism. *Journal of Autism and Developmental Disorders*, *43*(1), 114–121. doi: 10.1007/s10803-012-1550-8

White, S. J., Burgess, P. W., & Hill, E. L. (2009). Impairments on "open-ended" executive function tests in autism. *Autism Research*, *2*(3), 138–147. doi: 10.1002/aur.78

Whitson, J. A., & Galinsky, A. D. (2008). Lacking control increases illusory pattern perception. *Science*, *322*(5898), 115–117. doi: 10.1126/science.1159845

Wilson, T. D., & Gilbert, D. T. (2008). Explaining away: A model of affective adaptation. *Perspectives on Psychological Science*, *3*(5), 370–386. doi: 10.1111/j.1745-6924.2008.00085.x

Winkielman, P., & Cacioppo, J. T. (2001). Mind at ease puts a smile on the face: Psychophysiological evidence that processing facilitation elicits positive affect. *Journal of Personality and Social Psychology*, *81*(6), 989–1000. doi: 10.1037/0022-3514.81.6.989

Winkler, I., Denham, S. L., & Nelken, I. (2009). Modeling the auditory scene: predictive regularity representations and perceptual objects. *Trends in Cognitive Sciences*, *13*(12), 532–540. doi: 10.1016/j.tics.2009.09.003

Winstanley, C. A., Cocker, P. J., & Rogers, R. D. (2011). Dopamine modulates reward expectancy during performance of a slot machine task in rats: Evidence for a 'Near-miss' effect. *Neuropsychopharmacology*, *36*(5), 913–925. doi: 10.1038/npp.2010.230

Witkin, H. A., & Goodenough, D. R. (1981). Cognitive styles: essence

and origins. field dependence and field independence. *Psychological Issues*(51), 1–141.

Wolf, U., Rapoport, M. J., & Schweizer, T. A. (2009). Evaluating the affective component of the cerebellar cognitive affective syndrome. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *21*(3), 245–253. doi: 10.1176/appi.neuropsych.21.3.245

Wood, W., & Neal, D. T. (2007). A new look at habits and the habit-goal interface. *Psychological Review*, *114*(4), 843–863. doi: 10.1037/0033-295X.114.4.843

Wörgötter, F., & Porr, B. (2005). Temporal sequence learning, prediction, and control: a review of different models and their relation to biological mechanisms. *Neural Computation*, *17*(2), 245–319. doi: 10.1162/0899766053011555

Wuerger, S. M., Crocker-Buque, A., & Meyer, G. F. (2012). Evidence for auditory-visual processing specific to biological motion. *Seeing and perceiving*, *25*(1), 15–28. doi: 10.1163/187847611X620892

Wundt, W. (1897). *Outline of psychology* (Vol. xviii; C. H. Judd, Trans.). Leipzig, Germany: Wilhelm Engelmann.

Yang, E., Zald, D. H., & Blake, R. (2007). Fearful expressions gain preferential access to awareness during continuous flash suppression. *Emotion*, *7*(4), 882–886. doi: 10.1037/1528-3542.7.4.882

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, *8*(8), 665–670. doi: 10.1038/nmeth.1635

Yerkes, R., & Dodson, J. (1908). The relation of strength of stimulus to rapidity of habit formation. *Journal of Comparative Neurology & Psychology*, *18*, 459–482. doi: 10.1002/cne.920180503

Yu, A. J., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron*, *46*(4), 681–692. doi: 10.1016/j.neuron.2005.04.026

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *9*(2, Part 2), 1–27. doi: 10.1037/h0025848

Zaki, J. (2013). Cue integration a common framework for social cognition and physical perception. *Perspectives on Psychological Science*, *8*(3), 296–312. doi: 10.1177/1745691613475454

Zeki, S. (2004). The neurology of ambiguity. *Consciousness and Cognition*, *13*(1), 173–96. doi: 10.1016/j.concog.2003.10.003

Zwaigenbaum, L., Bryson, S., Rogers, T., Roberts, W., Brian, J., & Szatmari, P. (2005). Behavioral manifestations of autism in the first year of life. *International Journal of Developmental Neuro-*

science, *23*(2–3), 143–152. doi: 10.1016/j.ijdevneu.2004.05.001